

UNITED STATES AIR FORCE
SUMMER RESEARCH PROGRAM -- 1995
SUMMER RESEARCH EXTENSION PROGRAM FINAL REPORTS

VOLUME 3

ROME LABORATORY

RESEARCH & DEVELOPMENT LABORATORIES

5800 Uplander Way
Culver City, CA 90230-6608

Program Director, RDL
Gary Moore

Program Manager, AFOSR
Major David Hart

Program Manager, RDL
Scott Licoscas

Program Administrator, RDL
Gwendolyn Smith

Program Administrator, RDL
Johnetta Thompson

Submitted to:

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH
Bolling Air Force Base
Washington, D.C.

May 1996

20010319 024

AQM01-06-1069

REPORT DOCUMENTATION PAGE

AFRL-SR-BL-TR-00-

0698

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering the data, reviewing and collecting the information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to Washington Headquarters Service, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave blank)		2. REPORT DATE May, 1996		3. REPORT TYPE AND PERIOD	
4. TITLE AND SUBTITLE 1995 Summer Research Program (SRP), Summer Research Extension Program (SREP), Final Report, Volume 3, Rome Laboratory				5. FUNDING NUMBERS F49620-93-C-0063	
6. AUTHOR(S) Gary Moore					
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Research & Development Laboratories (RDL) 5800 Uplander Way Culver City, CA 90230-6608				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Air Force Office of Scientific Research (AFOSR) 801 N. Randolph St. Arlington, VA 22203-1977				10. SPONSORING/MONITORING AGENCY REPORT NUMBER	
11. SUPPLEMENTARY NOTES					
12a. DISTRIBUTION AVAILABILITY STATEMENT Approved for Public Release				12b. DISTRIBUTION CODE	
13. ABSTRACT (Maximum 200 words) The United States Air Force Summer Research Program (SRP) is designed to introduce university, college, and technical institute faculty members to Air Force research. This is accomplished by the faculty members, graduate students, and high school students being selected on a nationally advertised competitive basis during the summer intersession period to perform research at Air Force Research Laboratory (AFRL) Technical Directorates and Air Force Air Logistics Centers (ALC). AFOSR also offers its research associates (faculty only) an opportunity, under the Summer Research Extension Program (SREP), to continue their AFOSR-sponsored research at their home institutions through the award of research grants. This volume consists of the SREP program background, management information, statistics, a listing of the participants, and the technical report for each participant of the SREP working at the AF Rome Laboratory.					
14. SUBJECT TERMS Air Force Research, Air Force, Engineering, Laboratories, Reports, Summer, Universities, Faculty, Graduate Student, High School Student				15. NUMBER OF PAGES	
				16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT UL		

GENERAL INSTRUCTIONS FOR COMPLETING SF 298

The Report Documentation Page (RDP) is used in announcing and cataloging reports. It is important that this information be consistent with the rest of the report, particularly the cover and title page. Instructions for filling in each block of the form follow. It is important to **stay within the lines** to meet **optical scanning requirements**.

Block 1. Agency Use Only (*Leave blank*).

Block 2. Report Date. Full publication date including day, month, and year, if available
(e.g. 1 Jan 88). Must cite at least the year.

Block 3. Type of Report and Dates Covered. State whether report is interim, final, etc. If applicable, enter inclusive report dates (e.g. 10 Jun 87 - 30 Jun 88).

Block 4. Title and Subtitle. A title is taken from the part of the report that provides the most meaningful and complete information. When a report is prepared in more than one volume, repeat the primary title, add volume number, and include subtitle for the specific volume. On classified documents enter the title classification in parentheses.

Block 5. Funding Numbers. To include contract and grant numbers; may include program element number(s), project number(s), task number(s), and work unit number(s). Use the following labels:

C - Contract	PR - Project
G - Grant	TA - Task
PE - Program Element	WU - Work Unit Accession No.

Block 6. Author(s). Name(s) of person(s) responsible for writing the report, performing the research, or credited with the content of the report. If editor or compiler, this should follow the name(s).

Block 7. Performing Organization Name(s) and Address(es).
Self-explanatory.

Block 8. Performing Organization Report Number. Enter the unique alphanumeric report number(s) assigned by the organization performing the report.

Block 9. Sponsoring/Monitoring Agency Name(s) and Address(es).
Self-explanatory.

Block 10. Sponsoring/Monitoring Agency Report Number. (*If known*)

Block 11. Supplementary Notes. Enter information not included elsewhere such as: Prepared in cooperation with....; Trans. of....; To be published in.... When a report is revised, include a statement whether the new report supersedes or supplements the older report.

Block 12a. Distribution/Availability Statement. Denotes public availability or limitations. Cite any availability to the public. Enter additional limitations or special markings in all capitals (e.g. NOFORN, REL, ITAR).

DOD - See DoDD 5230.24, "Distribution Statements on Technical Documents."

DOE - See authorities.

NASA - See Handbook NHB 2200.2.

NTIS - Leave blank.

Block 12b. Distribution Code.

DOD - Leave blank.

DOE - Enter DOE distribution categories from the Standard Distribution for Unclassified Scientific and Technical Reports.

Leave blank.

NASA - Leave blank.

NTIS -

Block 13. Abstract. Include a brief (*Maximum 200 words*) factual summary of the most significant information contained in the report.

Block 14. Subject Terms. Keywords or phrases identifying major subjects in the report.

Block 15. Number of Pages. Enter the total number of pages.

Block 16. Price Code. Enter appropriate price code (*NTIS only*).

Blocks 17. - 19. Security Classifications. Self-explanatory. Enter U.S. Security Classification in accordance with U.S. Security Regulations (i.e., UNCLASSIFIED). If form contains classified information, stamp classification on the top and bottom of the page.

Block 20. Limitation of Abstract. This block must be completed to assign a limitation to the abstract. Enter either UL (unlimited) or SAR (same as report). An entry in this block is necessary if the abstract is to be limited. If blank, the abstract is assumed to be unlimited.

PREFACE

This volume is part of a five-volume set that summarizes the research of participants in the 1995 AFOSR Summer Research Extension Program (SREP). The current volume, Volume 1 of 5, presents the final reports of SREP participants at Armstrong Laboratory, Phillips Laboratory, Rome Laboratory, Wright Laboratory, Arnold Engineering Development Center, Frank J. Seiler Research Laboratory, and Wilford Hall Medical Center.

Reports presented in this volume are arranged alphabetically by author and are numbered consecutively -- e.g., 1-1, 1-2, 1-3; 2-1, 2-2, 2-3, with each series of reports preceded by a management summary. Reports in the five-volume set are organized as follows:

VOLUME	TITLE
1A	Armstrong Laboratory (part one)
1B	Armstrong Laboratory (part two)
2	Phillips Laboratory
3	Rome Laboratory
4A	Wright Laboratory (part one)
4B	Wright Laboratory (part two)
5	Arnold Engineering Development Center Frank J. Seiler Research Laboratory Wilford Hall Medical Center

1995 SREP FINAL REPORTS

Armstrong Laboratory

VOLUME 1

Report #	Report Title Author's University	Report Author
1	Determination of the Redox Capacity of Soil Sediment and Prediction of Pollutant University of Georgia, Athens, GA	Dr. James Anderson Analytical Chemistry AL/EQ
2	Finite Element Modeling of the Human Neck and Its Validation for the ATB Villanova University, Villanova, PA	Dr. Hashem Ashrafiuon Mechanical Engineering AL/CF
3	An Examination of the Validity of the Experimental Air Force ASVAB Composites Tulane University, New Orleans, LA	Dr. Michael Burke Psychology AL/HR
4	Fuel Identification by Neural Networks Analysis of the Response of Vapor Sensitive Sensors Arrays Edinboro University of Pennsylvania, Edinboro, PA	Dr. Paul Edwards Chemistry AL/EQ
5	A Comparison of Multistep vs Singlestep Arrhenius Integral Models for Describing Laser Induced Thermal Damage Florida International University, Miami, FL	Dr. Bernard Gerstman Physics AL/OE
6	Effects of Mental Workload and Electronic Support on Negotiation Performance University of Dayton, Dayton, OH	Dr. Kenneth Graetz Psychology AL/HR
7	Regression to the Mean in Half Life Studies University of Main, Orono, ME	Dr. Pushpa Gupta Mathematics & Statistics AL/AO
8	Application of the MT3D Solute Transport Model to the Made-2 Site: Calibration Florida State University, Tallahassee, FL	Dr. Manfred Koch Geophysics AL/EQ
9	Computer Calculations of Gas-Phase Reaction Rate Constants Florida State University, Tallahassee, FL	Dr. Mark Novotny SupercompComp. Res. I AL/EQ
10	Surface Fitting Three Dimensional Human Scan Data Ohio University, Athens, OH	Dr. Joseph Nurre Mechanical Engineering AL/CF
11	The Effects of Hyperbaric Oxygenation on Metabolism of Drugs and Other Xenobioti University of So. Carolina, Columbia, So. Carolina	Dr. Edward Piepmeier Pharmaceutics AL/AO
12	Maintaining Skills After Training: The Role of Opportunity to Perform Trained Tasks on Training Effectiveness Rice University, Houston, TX	Dr. Miguel Quinones Psychology AL/HR

1995 SREP FINAL REPORTS

Armstrong Laboratory

VOLUME 1 (cont.)

Report #	Report Title Author's University	Report Author
13	Nonlinear Transcutaneous Electrical Stimulation of the Vestibular System University of Illinois Urbana-Champaign, Urbana, IL	Dr. Gary Riccio Psychology AL/CF
14	Documentation of Separating and Separated Boundary Layer Flow, For Application Texas A&M University, College Station, TX	Dr. Wayne Shebilske Psychology AL/HR
15	Tactile Feedback for Simulation of Object Shape and Textural Information in Haptic Displays Ohio State University, Columbus, OH	Dr. Janet Weisenberger Speech & Hearing AL/CF
16	Melatonin Induced Prophylactic Sleep as a Countermeasure for Sleep Deprivation Oregon Health Sciences University, Portland, OR	Mr. Rod Hughes Psychology AL/CF

1995 SREP FINAL REPORTS

Phillips Laboratory

VOLUME 2A

Report #	Report Title Author's University	Report Author
1	Investigation of the Mixed-Mode Fracture Behavior of Solid Propellants University of Houston, Houston, TX	Dr. K. Ravi-Chandar Aeronautics PL/RK
2	Performance Study of ATM-Satellite Network SUNY-Buffalo, Buffalo, NY	Dr. Nasser Ashgriz Mechanical Engineering PL/RK
3	Characterization of CMOS Circuits Using a Highly Calibrated Low-Energy X-Ray Source Embry-Riddle Aeronautical Univ., Prescott, AZ	Dr. Raymond Bellem Computer Science PL/VT
4	Neutron Diagnostics for Pulsed Plasmas of Compact Toroid-Marauder Type Stevens Institute of Tech, Hoboken, NJ	Dr. Jan Brzosko Nuclear Physics PL/WS
5	Parallel Computation of Zernike Aberration Coefficients for Optical Aberration Correction University of Houston-Victoria, Victoria, TX	Dr. Meledath Damodaran Math & Computer Science PL/LI
6	Quality Factor Evaluation of Complex Cavities University of Denver, Denver, CO	Dr. Ronald DeLyser Electrical Engineering PL/WS
7	Unidirectional Ring Lasers and Laser Gyros with Multiple Quantum Well Gain University of New Mexico, Albuquerque, NM	Dr. Jean-Claude Diels Physics PL/LI
8	A Tool for the Formation of Variable Parameter Inverse Synthetic Aperture Radar University of Nevada, Reno, NV	Dr. James Henson Electrical Engineering PL/WS
9	Radar Ambiguity Functionals Univ. of Massachusetts at Lowell, Lowell, MA	Dr. Gerald Kaiser Physics PL/GP
10	The Synthesis and Chemistry of Peroxonitrites Peroxonitrous Acid Univ. of Massachusetts at Lowell, Lowell, MA	Dr. Albert Kowalak Chemistry PL/GP
11	Temperature and Pressure Dependence of the Band Gaps and Band Offsets University of Houston, Houston, TX	Dr. Kevin Malloy Electrical Engineering PL/VT
12	Theoretical Studies of the Performance of Novel Fiber-Coupled Imaging Interferom University of New Mexico, Albuquerque, NM	Dr. Sudhakar Prasad Physics PL/LI

1995 SREP FINAL REPORTS

Phillips Laboratory

VOLUME 2B

Report #	Author's University	Report Author
13	Static and Dynamic Graph Embedding for Parallel Programming Texas AandM Univ.-Kingsville, Kingsville, TX	Dr. Mark Purtill Mathematics PL/WS
14	Ultrafast Process and Modulation in Iodine Lasers University of New Mexico, Albuquerque, NM	Dr. W. Rudolph Physics PL/LI
15	Impedance Matching and Reflection Minimization for Transient EM Pulses Through University of New Mexico, Albuquerque, NM	Dr. Alexander Stone Mathematics and Statics PL/WS
16	Low Power Retromodular Based Optical Transceiver for Satellite Communications Utah State University, Logan, UT	Dr. Charles Swenson Electrical Engineering PL/VT
17	Improved Methods of Tilt Measurement for Extended Images in the Presence of Atmospheric Disturbances Using Optical Flow Michigan Technological Univ., Houghton, MI	Mr. John Lipp Electrical Engineering PL/LI
18	Thermoluminescence of Simple Species in Molecular Hydrogen Matrices Cal State Univ.-Northridge, Northridge, CA	Ms. Janet Petroski Chemistry PL/RK
19	Design, Fabrication, Intelligent Cure, Testing, and Flight Qualification University of Cincinnati, Cincinnati, OH	Mr. Richard Salasovich Mechanical Engineering PL/VT

1995 SREP FINAL REPORTS

Rome Laboratory

VOLUME 3

Report #	Author's University	Report Author
1	Performance Study of an ATM/Satellite Network Florida Atlantic University, Boca Raton, FL	Dr. Valentine Aalo Electrical Engineering RL/C3
2	Interference Excision in Spread Spectrum Communication Systems Using Time-Frequency Distributions Villanova University, Villanova, PA	Dr. Moeness Amin Electrical Engineering RL/C3
3	Designing Software by Reformulation Using KIDS Oklahoma State University, Stillwater, OK	Dr. David Benjamin Computer Science RL/C3
4	Detection Performance of Over Resolved Targets with Non-Uniform and Non-Gaussian Howard University, Washington, DC	Dr. Ajit Choudhury Engineering RL/OC
5	Computer-Aided-Design Program for Solderless Coupling Between Microstrip and Stripline Structures Southern Illinois University, Carbondale, IL	Dr. Frances Harackiewicz Electrical Engineering RL/ER
6	Spanish Dialect Identification Project Colorado State University, Fort Collins, CO	Dr. Beth Losiewicz Psycholinguistics RL/IR
7	Automatic Image Registration Using Digital Terrain Elevation Data University of Maine, Orono, ME	Dr. Mohamed Musavi Engineering RL/IR
8	Infrared Images of Electromagnetic Fields University of Colorado, Colorado Springs, CO	Dr. John Norgard Engineering RL/ER
9	Femtosecond Pump-Probe Spectroscopy System SUNY Institute of Technology, Utica, NY	Dr. Dean Richardson Photonics RL/OC
10	Synthesis and Properties B-Diketonate-Modified Heterobimetallic Alkoxides Tufts University, Medford, MA	Dr. Daniel Ryder, Jr. Chemical Engineering RL/ER
11	Optoelectronic Study of Semiconductor Surfaces and Interfaces Rensselaer Polytechnic Institute, Troy, NY	Dr. Xi-Cheng Zhang Physics RL/ER

1995 SREP FINAL REPORTS

Wright Laboratory

VOLUME 4A

Report #	Author's University	Report Author
1	An Investigation of the Heating and Temperature Distribution in Electrically Excited Foils Auburn University, Auburn, AL	Dr. Michael Baginski Electrical Engineering WL/MN
2	Micromechanics of Creep in Metals and Ceramics at High Temperature Wayne State University, Detroit, MI	Dr. Victor Berdichevsky Aerospace Engineering WL/FI
3	Development of a Fluorescence-Based Chemical Sensor for Simultaneous Oxygen Quantitation and Temp. Measurement Columbus College, Columbus, GA	Dr. Steven Buckner Chemistry WL/PO
4	Development of High-Performance Active Dynamometer Sys. for Machines and Drive Clarkson University, Potsdam, NY	Dr. James Carroll Electrical Engineering WL/PO
5	SOLVING $z(t)=1n[Acos(w_1t)+Bcos(w_2)+C]$ Transylvania University, Lexington, KY	Dr. David Choate Mathematics WL/AA
6	Synthesis, Processing and Characterization of Nonlinear Optical Polymer Thin Films University of Cincinnati, Cincinnati, OH	Dr. Stephen Clarson Mats Science & Engineering WL/ML
7	An Investigation of Planning and Scheduling Algorithms for Sensor Management Embry-Riddle Aeronautical University, Prescott, AZ	Dr. Milton Cone Comp. Science & Engineering WL/AA
8	A Study to Determine Wave Gun Firing Cycles for High Performance Model Launches Louisiana State University, Baton Rouge, LA	Dr. Robert Courter Mechanical Engineering WL/MN
9	Characterization of Electro-Optic Polymers University of Dayton, Dayton, OH	Dr. Vincent Dominic Electro Optics Program WL/ML
10	A Methodology for Affordability in the Design Process Clemson University, Clemson, SC	Dr. Georges Fadel Mechanical Engineering WL/MT
11	Data Reduction and Analysis for Laser Doppler Velocimetry North Carolina State University, Raleigh, NC	Dr. Richard Gould Mechanical Engineering WL/PO

1995 SREP FINAL REPORTS

Wright Laboratory

VOLUME 4A (cont.)

Report #	Author's University	Report Author
12	Hyperspectral Target Identification Using Bomen Spectrometer Data University of Dayton, Dayton, OH	Dr. Russell Hardie Electrical Engineering WL/AA
13	Robust Fault Detection and Classification Auburn University, Auburn, AL	Dr. Alan Hodel Electrical Engineering WL/MN
14	Multidimensional Algorithm Development and Analysis Mississippi State University, Mississippi State University, MS	Dr. Jonathan Janus Aerospace Engineering WL/MN
15	Characterization of Interfaces in Metal-Matrix Composites Michigan State University, East Lansing, MI	Dr. Iwona Jasiuk Materials Science WL/ML
16	TSI Mitigation: A Mountaintop Database Study Lafayette College, Easton, PA	Dr. Ismail Jouny Electrical Engineering WL/AA
17	Comparative Study and Performance Analysis of High Resolution SAR Imaging Techniques University of Florida, Gainesville, FL	Dr. Jian Li Electrical Engineering WL/AA

1995 SREP FINAL REPORTS

Wright Laboratory

VOLUME 4B

Report #	Author's University	Report Author
18	Prediction of Missile Trajectory University of Missouri-Columbia, Columbia, MO	Dr. Chun-Shin Lin Electrical Engineering WL/FI
19	Three Dimensional Deformation Comparison Between Bias and Radial Aircraft Tires Cleveland State University, Cleveland, OH	Dr. Paul Lin Mechanical Engineering WL/FI
20	Investigation of AlGaAs/GaAs Heterojunctin Bipolar Transistor Reliability Based University of Central Florida, Orlando, FL	Dr. Juin Liou Electrical Engineering WL/EL
21	Thermophysical Invariants From LWIR Imagery for ATR University of Virginia, Charlottesville, VA	Dr. Nagaraj Nandhakumar Electrical Engineering WL/AA
22	Effect of Electromagnetic Environment on Array Signal Processing University of Dayton, Dayton, OH	Dr. Krishna Pasala Electrical Engineering WL/AA
23	Functional Decomposition of Binary, Multiple-Valued, and Fuzzy Logic Portland State University, Portland, OR	Dr. Marek Perkowski Electrical Engineering WL/AA
24	Superresolution of Passive Millimeter-Wave Imaging Auburn University, Auburn, AL	Dr. Stanley Reeves Electrical Engineering WL/MN
25	Development of a Penetrator Optimizer University of Alabama, Tuscaloosa, AL	Dr. William Rule Engineering Science WL/MN
26	Heat Transfer for Turbine Blade Film Cooling with Free Stream Turbulence-Measurements and Predictions University of Dayton, Dayton, OH	Dr. John Schauer Mech. & Aerosp. Engineering WL/FI
27	Neural Network Identification and Control in Metal Forging University of Florida, Gainesville, FL	Dr. Carla Schwartz Electrical Engineering WL/FI
28	Documentation of Separating and Separated Boundary Layer Flow, for Application University of Minnesota, Minneapolis, MN	Dr. Terrence Simon Mechanical Engineering WL/PO
29	Transmission Electron Microscopy of Semiconductor Heterojunctions Carnegie Melon University, Pittsburgh, PA	Dr. Marek Skowronski Matls Science & Engineering WL/EL

1995 SREP FINAL REPORTS

Wright Laboratory

VOLUME 4B (cont.)

Report #	Author's University	Report Author
30	Parser in SWI-PROLOG Wright State University, Dayton, OH	Dr. K. Thirunarayan Computer Science WL/EL
31	Development of Qualitative Process Control Discovery Systems for Polymer Composite and Biological Materials University of California, Los Angeles, CA	Dr. Robert Trelease Anatomy & Cell Biology WL/ML
32	Improved Algorithm Development of Massively Parallel Epic Hydrocode in Cray T3D Massively Parallel Computer Florida Atlantic University, Boca Raton, FL	Dr. Chi-Tay Tsai Engineering Mechanics WL/MN
33	The Characterization of the Mechanical Properties of Materials in a Biaxial Stress Environment University of Kentucky, Lexington, KY	Dr. John Lewis Materials Science Engineering WL/MN

1995 SREP FINAL REPORTS

VOLUME 5

Report #	Author's University	Report Author
Arnold Engineering Development Center		
1	Plant-Wide Preventive Maintenance and Monitoring Vanderbilt University	Mr. Theodore Bapty Electrical Engineering AEDC
Frank J. Seiler Research Laboratory		
1	Block Copolymers at Inorganic Solid Surfaces Colorado School of Mines, Golden, CO	Dr. John Dorgan Chemical Engineering FJSRL
2	Non-Linear Optical Properties of Polyacetylenes and Related Barry University, Miami, FL	Dr. M. A. Jungbauer Chemistry FJSRL
3	Studies of Second Harmonic Generation in Glass Waveguides Allegheny College, Meadville, PA	Dr. David Statman Physics FJSRL
Wilford Hall Medical Center		
1	Biochemical & Cell Physiological Aspects of Hyperthermia University of Miami, Coral Gables, FL	Dr. W. Drost-Hansen Chemistry WHMC

1995 SUMMER RESEARCH EXTENSION PROGRAM (SREP) MANAGEMENT REPORT

1.0 BACKGROUND

Under the provisions of Air Force Office of Scientific Research (AFOSR) contract F49620-90-C-0076, September 1990, Research & Development Laboratories (RDL), an 8(a) contractor in Culver City, CA, manages AFOSR's Summer Research Program. This report is issued in partial fulfillment of that contract (CLIN 0003AC).

The Summer Research Extension Program (SREP) is one of four programs AFOSR manages under the Summer Research Program. The Summer Faculty Research Program (SFRP) and the Graduate Student Research Program (GSRP) place college-level research associates in Air Force research laboratories around the United States for 8 to 12 weeks of research with Air Force scientists. The High School Apprenticeship Program (HSAP) is the fourth element of the Summer Research Program, allowing promising mathematics and science students to spend two months of their summer vacations working at Air Force laboratories within commuting distance from their homes.

SFRP associates and exceptional GSRP associates are encouraged, at the end of their summer tours, to write proposals to extend their summer research during the following calendar year at their home institutions. AFOSR provides funds adequate to pay for SREP subcontracts. In addition, AFOSR has traditionally provided further funding, when available, to pay for additional SREP proposals, including those submitted by associates from Historically Black Colleges and Universities (HBCUs) and Minority Institutions (MIs). Finally, laboratories may transfer internal funds to AFOSR to fund additional SREPs. Ultimately the laboratories inform RDL of their SREP choices, RDL gets AFOSR approval, and RDL forwards a subcontract to the institution where the SREP associate is employed. The subcontract (see Appendix 1 for a sample) cites the SREP associate as the principal investigator and requires submission of a report at the end of the subcontract period.

Institutions are encouraged to share costs of the SREP research, and many do so. The most common cost-sharing arrangement is reduction in the overhead, fringes, or administrative charges institutions would normally add on to the principal investigator's or research associate's labor. Some institutions also provide other support (e.g., computer run time, administrative assistance, facilities and equipment or research assistants) at reduced or no cost.

When RDL receives the signed subcontract, we fund the effort initially by providing 90% of the subcontract amount to the institution (normally \$18,000 for a \$20,000 SREP). When we receive the end-of-research report, we evaluate it administratively and send a copy to the laboratory for a technical evaluation. When the laboratory notifies us the SREP report is acceptable, we release the remaining funds to the institution.

2.0 THE 1995 SREP PROGRAM

SELECTION DATA: A total of 719 faculty members (SFRP Associates) and 286 graduate students (GSRP associates) applied to participate in the 1994 Summer Research Program. From these applicants 185 SFRPs and 121 GSRPs were selected. The education level of those selected was as follows:

1994 SRP Associates, by Degree			
SFRP		GSRP	
PHD	MS	MS	BS
179	6	52	69

Of the participants in the 1994 Summer Research Program 90 percent of SFRPs and 25 percent of GSRPs submitted proposals for the SREP. Ninety proposals from SFRPs and ten from GSRPs were selected for funding, which equates to a selection rate of 54% of the SFRP proposals and of 34% for GSRP proposals.

1995 SREP: Proposals Submitted vs. Proposals Selected			
	Summer 1994 Participants	Submitted SREP Proposals	SREPs Funded
SFRP	185	167	90
GSRP	121	29	10
TOTAL	306	196	100

The funding was provided as follows:

Contractual slots funded by AFOSR	75
Laboratory funded	14
Additional funding from AFOSR	<u>11</u>
Total	100

Six HBCU/MI associates from the 1994 summer program submitted SREP proposals; six were selected (none were lab-funded; all were funded by additional AFOSR funds).

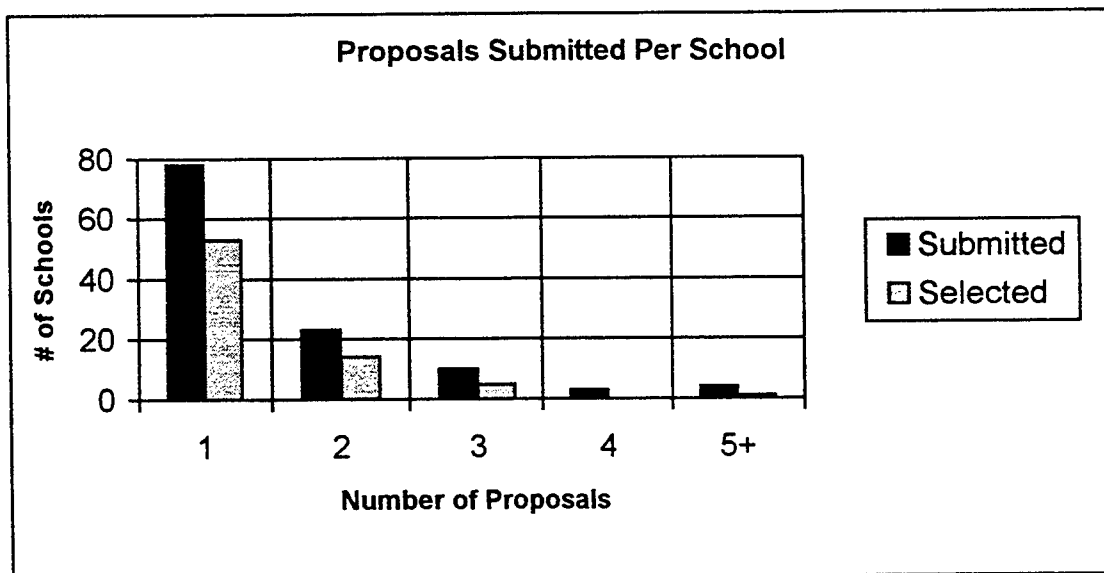
Proposals Submitted and Selected, by Laboratory		
	Applied	Selected
Armstrong Laboratory	41	19
Arnold Engineering Development Center	12	4
Frank J. Seiler Research Laboratory	6	3
Phillips Laboratory	33	19
Rome Laboratory	31	13
Wilford Hall Medical Center	2	1
Wright Laboratory	62	37
TOTAL		

Note: Phillips Laboratory funded 3 SREPs; Wright Laboratory funded 11; and AFOSR funded 11 beyond its contractual 75.

The 306 1994 Summer Research Program participants represented 135 institutions.

Institutions Represented on the 1994 SRP and 1995 SREP		
Number of schools represented in the Summer 92 Program	Number of schools represented in submitted proposals	Number of schools represented in Funded Proposals
135	118	73

Forty schools had more than one participant submitting proposals.



The selection rate for the 78 schools submitting 1 proposal (68%) was better than those submitting 2 proposals (61%), 3 proposals (50%), 4 proposals (0%) or 5+ proposals (25%). The 4 schools that submitted 5+ proposals accounted for 30 (15%) of the 196 proposals submitted.

Of the 196 proposals submitted, 159 offered institution cost sharing. Of the funded proposals which offered cost sharing, the minimum cost share was \$1000.00, the maximum was \$68,000.00 with an average cost share of \$12,016.00.

Proposals and Institution Cost Sharing		
	Proposals Submitted	Proposals Funded
With cost sharing	159	82
Without cost sharing	37	18
Total	196	100

The SREP participants were residents of 41 different states. Number of states represented at each laboratory were:

States Represented, by Proposals Submitted/Selected per Laboratory		
	Proposals Submitted	Proposals Funded
Armstrong Laboratory	21	13
Arnold Engineering Development Center	5	2
Frank J. Seiler Research Laboratory	5	3
Phillips Laboratory	16	14
Rome Laboratory	14	7
Wilford Hall Medical Center	2	1
Wright Laboratory	24	20

Eleven of the 1995 SREP Principal Investigators also participated in the 1994 SREP.

ADMINISTRATIVE EVALUATION: The administrative quality of the SREP associates' final reports was satisfactory. Most complied with the formatting and other instructions provided to them by RDL. Ninety seven final reports and two interim reports have been received and are included in this report. The subcontracts were funded by \$1,991,623.00 of Air Force money. Institution cost sharing totaled \$985,353.00.

TECHNICAL EVALUATION: The form used for the technical evaluation is provided as Appendix 2. ninety-two evaluation reports were received. Participants by laboratory versus evaluations submitted is shown below:

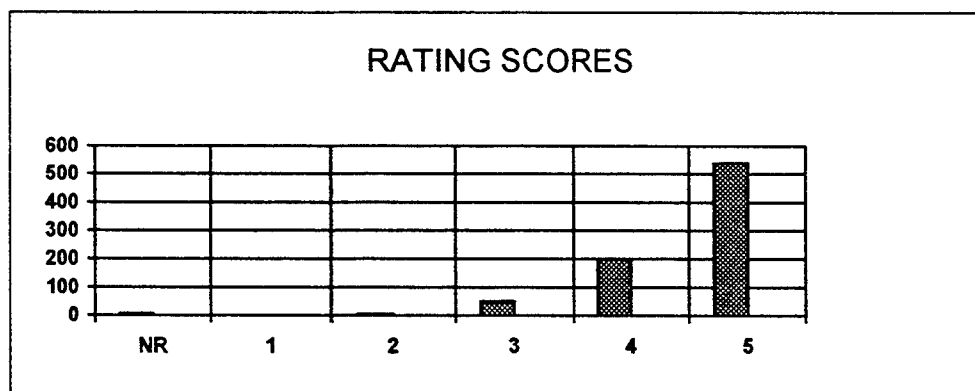
	Participants	Evaluations	Percent
Armstrong Laboratory	23 ¹	20	95.2
Arnold Engineering Development Center	4	4	100
Frank J. Seiler Research Laboratory	3	3	100
Phillips Laboratory	19 ²	18	100
Rome Laboratory	13	13	100
Wilford Hall Medical Center	1	1	100
Wright Laboratory	37	34	91.9
Total			

Notes:

- 1: Research on two of the final reports was incomplete as of press time so there aren't any technical evaluations on them to process, yet. Percent complete is based upon 20/21=95.2%
- 2: One technical evaluation was not completed because one of the final reports was incomplete as of press time. Percent complete is based upon 18/18=100%
- 3: See notes 1 and 2 above. Percent complete is based upon 93/97=95.9%

The number of evaluations submitted for the 1995 SREP (95.9%) shows a marked improvement over the 1994 SREP submittals (65%).

PROGRAM EVALUATION: Each laboratory focal point evaluated ten areas (see Appendix 2) with a rating from one (lowest) to five (highest). The distribution of ratings was as follows:



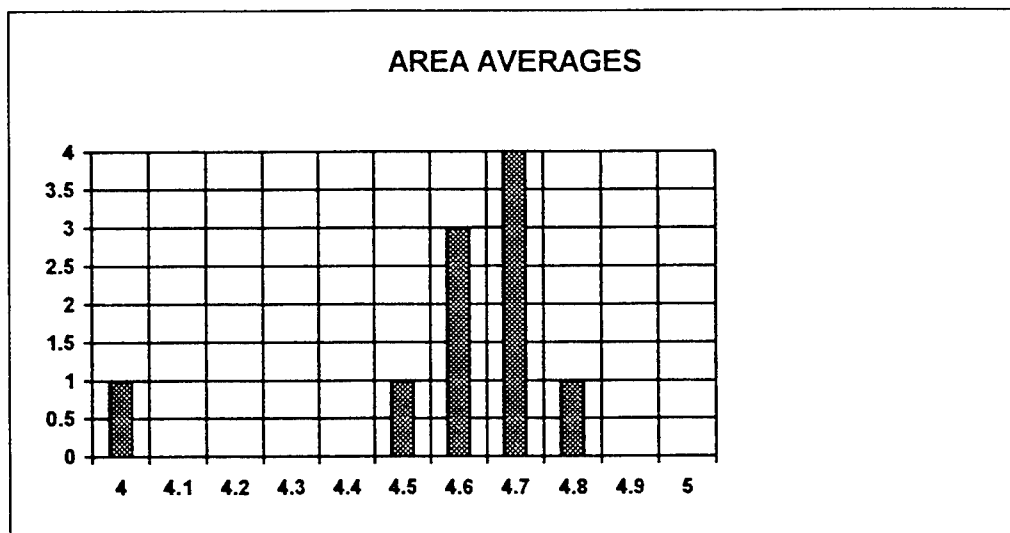
Rating	Not Rated	1	2	3	4	5
# Responses	7	1	7	62 (6%)	226 (25%)	617 (67%)

The 8 low ratings (one 1 and seven 2's) were for question 5 (one 2) "The USAF should continue to pursue the research in this SREP report" and question 10 (one 1 and six 2's) "The

one-year period for complete SREP research is about right”, in addition over 30% of the threes (20 of 62) were for question ten. The average rating by question was:

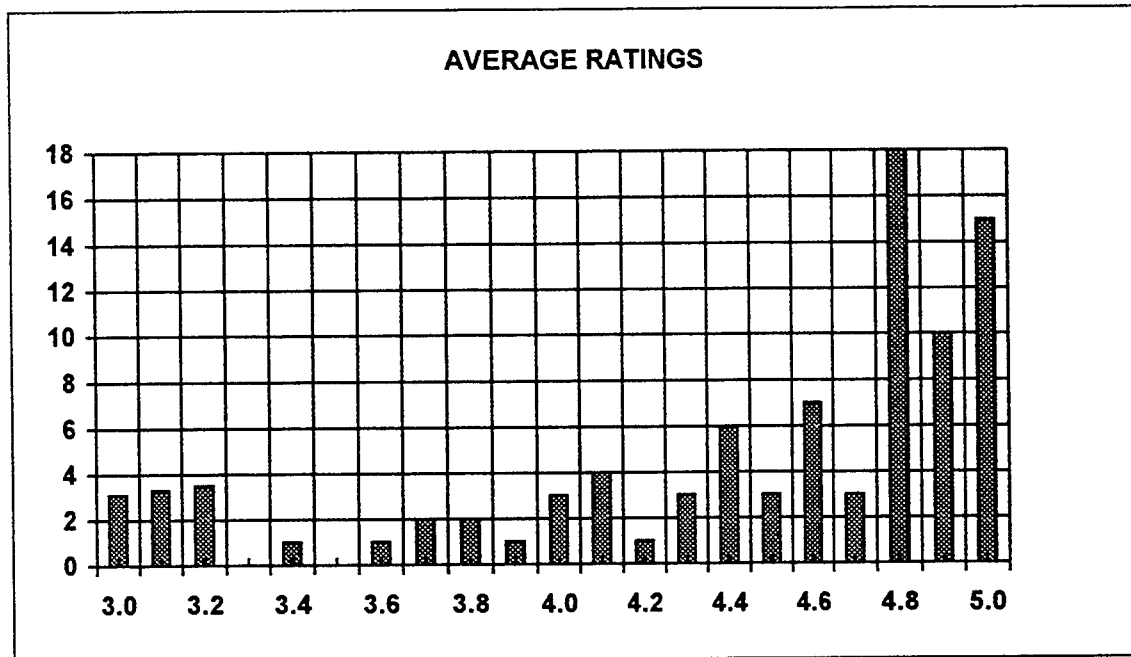
Question	1	2	3	4	5	6	7	8	9	10
Average	4.6	4.6	4.7	4.7	4.6	4.7	4.8	4.5	4.6	4.0

The distribution of the averages was:



Area 10 “the one-year period for complete SREP research is about right” had the lowest average rating (4.1). The overall average across all factors was 4.6 with a small sample standard deviation of 0.2. The average rating for area 10 (4.1) is approximately three sigma lower than the overall average (4.6) indicating that a significant number of the evaluators feel that a period of other than one year should be available for complete SREP research.

The average ratings ranged from 3.4 to 5.0. The overall average for those reports that were evaluated was 4.6. Since the distribution of the ratings is not a normal distribution the average of 4.6 is misleading. In fact over half of the reports received an average rating of 4.8 or higher. The distribution of the average report ratings is as shown:



It is clear from the high ratings that the laboratories place a high value on AFOSR's Summer Research Extension Programs.

3.0 SUBCONTRACTS SUMMARY

Table 1 provides a summary of the SREP subcontracts. The individual reports are published in volumes as shown:

<u>Laboratory</u>	<u>Volume</u>
Armstrong Laboratory	1A, 1B
Arnold Engineering Development Center	5
Frank J. Seiler Research Laboratory	5
Phillips Laboratory	2
Rome Laboratory	3
Wilford Hall Medical Center	5
Wright Laboratory	4A, 4B

1995 SREP SUB-CONTRACT DATA

Report Author Author's University	Author's Degree	Sponsoring Lab	Performance Period	Contract Amount	Univ. Cost Share
Anderson , James Analytical Chemistry University of Georgia, Athens, GA	PhD 95-0807	AL/EQ	01/01/95 12/31/95	\$25000.00	\$1826.00
			Determination of the Redox Capacity of Soil Sediment and Prediction of Pollutant		
Ashrafiuon , Hashem Mechanical Engineering Villanova University, Villanova, PA	PhD 95-0800	AL/CF	01/01/95 12/31/95	\$25000.00	\$19528.00
			Finite Element Modeling of the Human Neck and Its Validation for the ATB Model		
Burke , Michael Tulane University Tulane University, New Orleans, LA	PhD 95-0811	AL/HR	01/01/95 09/30/95	\$25000.00	\$1818.00
			An Examination of the Validity of the New Air Force ASVAB Composites		
Edwards , Paul Chemistry Edinboro Univ of Pennsylvania, Edinboro, PA	PhD 95-0808	AL/EQ	01/01/95 12/31/95	\$25000.00	\$5000.00
			Fuel Identification by Neural Networks Analysis of the Response of Vapor Sensiti		
Gerstman , Bernard Physics Florida International Universi, Miami, FL	PhD 95-0815	AL/OE	01/01/95 12/31/95	\$24289.00	\$2874.00
			A Comparison of Multistep vs Singlestep Arrhenius Integral Models for Describing		
Graetz , Kenneth Department of Psychology University of Dayton, Dayton, OH	PhD 95-0812	AL/HR	01/01/95 12/31/95	\$25000.00	\$0.00
			Effects of Mental Workload and Electronic Support on Negotiation Performance		
Gupta , Pushpa Mathematics University of Maine, Orono, ME	PhD 95-0802	AL/AO	01/01/95 12/31/95	\$25000.00	\$2859.00
			Regression to the Mean in Half Life Studies		
Koch , Manfred Geophysics Florida State University, Tallahassee, FL	PhD 95-0809	AL/EQ	12/01/94 04/30/95	\$25000.00	\$0.00
			Application of the MT3D Solute Transport Model to the Made-2 Site: Calibration		
Novotny , Mark Supercomputer Comp Res. I Florida State University, Tallahassee, FL	PhD 95-0810	AL/EQ	01/01/95 12/31/95	\$25000.00	\$0.00
			Computer Calculations of Gas-Phase Reaction Rate Constants		
Nurre , Joseph Mechanical Engineering Ohio University, Athens, OH	PhD 95-0804	AL/CF	01/01/95 12/31/95	\$25000.00	\$20550.00
			Surface Fitting Three Dimensional Human Head Scan Data		
Piepmeier , Edward Pharmaceutics University of South Carolina, Columbia, SC	PhD 95-0801	AL/AO	01/01/95 12/31/95	\$25000.00	\$11740.00
			The Effects of Hyperbaric Oxygenation on Metabolism of Drugs and Other Xenobioti		
Quinones , Miguel Psychology Rice University, Houston, TX	PhD 95-0813	AL/HR	01/01/95 12/31/95	\$25000.00	\$4000.00
			Maintaining Skills After Training: The Role of Opportunity to Perform Trained T		
Riccio , Gary Psychology Univ of IL Urbana-Champaign, Urbana, IL	PhD 95-0806	AL/CF	01/01/95 05/31/95	\$22931.00	\$0.00
			Nonlinear Transcutaneous Electrical Stimulation of the Vestibular System		
Shebilske , Wayne Dept of Psychology Texas A&M University, College Station, TX	PhD 95-0814	AL/HR	01/01/95 12/31/95	\$25000.00	\$5614.00
			Cognitive Factors in Distr Training Effects During Acquisition of Complex Skills		

1995 SREP SUB-CONTRACT DATA

Report Author Author's University	Author's Degree	Sponsoring Lab	Performance Period	Contract Amount	Univ. Cost Share
Weisenberger , Janet Dept of Speech & Hearing Ohio State University, Columbus, OH	PhD 95-0805	AL/CF	01/01/95 12/31/95	\$25000.00	\$12234.00
			Tactile Feedback for Simulation of Object Shape and Textural Information in Hapt		
Hughes , Rod Psychology Oregon Health Sciences University, Portland, OR	MA 95-0803	AL/CF	01/01/95 12/31/95	\$25000.00	\$0.00
			Melatonin Induced Prophylactic Sleep as a Countermeasure for Sleep Deprivation		
Bapty , Theodore Electrical Engineering Vanderbilt University, Nashville, TN	MS 95-0848	AEDC/E	01/01/95 12/31/95	\$24979.00	\$0.00
			Plant-Wide Preventive Maintenance & Monitoring		
Dorgan , John Chemical Engineering Colorado School of Mines, Golden, CO	PhD 95-0834	FJSRL/F	01/01/95 12/31/95	\$25000.00	\$0.00
			Block Copolymers at Inorganic Solid Surfaces		
Jungbauer , Mary Ann Chemistry Barry University, Miami, FL	PhD 95-0836	FJSRL/F	01/01/95 12/31/95	\$25000.00	\$24714.00
			Non-Linear Optical Properties of Polyacetylenes and Related Substituted Compound		
Statman , David Physics Allegheny College, Meadville, PA	PhD 95-0835	FJSRL/F	01/01/95 12/31/95	\$25000.00	\$6500.00
			Studies of Second Harmonic Generation in Glass Waveguides		
, Krishnaswamy Aeronautics University of Houston, Houston, TX	PhD 95-0818	PL/RK	01/01/95 12/31/95	\$24993.00	\$8969.00
			Mixed-Mode Fracture of Solid Propellants		
Ashgriz , Nasser Mechanical Engineering SUNY-Buffalo, Buffalo, NY	PhD 95-0816	PL/RK	01/01/95 12/31/95	\$25000.00	\$22329.00
			Effects of the Jet Characteristics on the Atomization and Mixing in A Pair of Im		
Bellem , Raymond Computer Science Embry-Riddle Aeronautical Univ, Prescott, AZ	PhD 95-0817	PL/VT	12/01/94 11/30/95	\$20000.00	\$8293.00
			Experimental Studies of the Effects of Ionizing Radiation on Commerically Proces		
Brzosko , Jan Nuclear Physics Stevens Institute of Tech, Hoboken, NJ	PhD 95-0828	PL/WS	11/01/94 02/01/95	\$24943.00	\$0.00
			Neutron Diagnostics for Pulsed Plasmas of Compact Toroid - Marauder Type		
Damodaran , Meledath Math & Computer Science University of Houston-Victoria, Victoria, TX	PhD 95-0831	PL/LI	01/01/95 12/31/94	\$24989.00	\$9850.00
			Parallel Computation of Zernike Aberration Coefficients for Optical Aber Correct		
DeLyser , Ronald Electrical Engineering University of Denver, Denver, CO	PhD 95-0877	PL/WS	01/01/95 12/31/95	\$25000.00	\$46066.00
			Quality Factor Evaluation of Complex Cavities		
Diels , Jean-Claude Physics University of New Mexico, Albuquerque, NM	PhD 95-0819	PL/LI	01/01/95 12/31/95	\$25000.00	\$0.00
			Unidirectional Ring Lasers and Laseer Gyros with Multiple Quantum Well Gain Medi		
Henson , James Electrical Engineering University of Nevada, Reno, NV	PhD 95-0820	PL/WS	01/01/95 12/31/95	\$25000.00	\$0.00
			Automatic Feature Extraction and Assessment of Wideband Range-Doppler Imagery of		
Kaiser , Gerald Physics University of Mass/Lowell, Lowell, MA	PhD 95-0821	PL/GP	01/01/95 12/31/95	\$25000.00	\$5041.00
			Multiresolution Analysis with Physical Wavelets		

1995 SREP SUB-CONTRACT DATA

Report Author Author's University	Author's Degree	Sponsoring Lab	Performance Period	Contract Amount	Univ. Cost Share
Kowalak , Albert Chemistry University of Massachusetts/Lo, Lowell, MA	PhD 95-0822	PL/GP The Synthesis and Chemistry of Peroxonitrites and Peroxonitrous Acid	01/01/95 12/31/95	\$24996.00	\$4038.00
Malloy , Kevin Electrical Engineering University of New Mexico, Albuquerque, NM	PhD 95-0829	PL/VT Temperature & Pressure Dependence of the Band Gaps & Band Offsets	01/01/95 12/31/95	\$24999.00	\$0.00
Prasad , Sudhakar Physics University of New Mexico, Albuquerque, NM	PhD 95-0823	PL/LI Theoretical Studies of the Performance of Novel Fiber-Coupled Imaging Interferom	01/01/95 12/31/95	\$25000.00	\$11047.00
Purtill , Mark Mathematics Texas A&M Univ-Kingsville, Kingsville, TX	PhD 95-0824	PL/WS Static and Dynamic Graph Embedding for Parallel Programming	01/01/95 12/31/95	\$25000.00	\$100.00
Rudolph , Wolfgang Physics University of New Mexico, Albuquerque, NM	PhD 95-0833	PL/LI Ultrafast Process and Modulation in Iodine Lasers	01/01/95 12/31/95	\$24982.00	\$6000.00
Stone , Alexander Mathematics & Statistics University of New Mexico, Alburquerque, NM	PhD 95-0827	PL/WS Impedance Matching And Reflection Minimization For Transient EM Pulses Through D	01/01/95 12/31/95	\$24969.00	\$0.00
Swenson , Charles Dept of Electrical Engr Utah State University, Logan, UT	PhD 95-0826	PL/VT Low Power Retromodulator based Optical Transceiver for Satellite Communications	01/01/95 12/31/95	\$25000.00	\$25000.00
Lipp , John Electrical Engineering Michigan Technological Univ, Houghton, MI	MS 95-0832	PL/LI Improved Methods of Tilt Measurement for Extended Images in the Presence of Atmo	01/01/95 12/31/95	\$24340.00	\$15200.00
Petroski , Janet Chemistry Cal State Univ/Northridge, Northridge, CA	BA 95-0830	PL/RK Thermoluminescence of Simple Species in Molecular Hydrogen Matrices	10/01/94 12/31/94	\$4279.00	\$0.00
Salasovich , Richard Mechanical Engineering University of Cincinnati, Cincinnati, OH	MS 95-0825	PL/VT Design, Fabrication, Intelligent Cure, Testing, and Flight Qualification of an A	01/01/95 12/31/95	\$25000.00	\$4094.00
Aalo , Valentine Dept of Electrical Engr Florida Atlantic University, Boca Raton, FL	PhD 95-0837	RL/C3 Performance Study of an ATM/Satellite Network	01/01/95 12/31/95	\$25000.00	\$13120.00
Amin , Moeness Electrical Engineering Villanova University, Villanova, PA	PhD 95-0838	RL/C3 Interference Excision in Spread Spectrum Communication Systems Using Time-Freque	01/01/95 12/31/95	\$25000.00	\$34000.00
Benjamin , David Computer Science Oklahoma State University, Stillwater, OK	PhD 95-0839	RL/C3 Designing Software by Decomposition using KIDS	01/01/95 12/31/95	\$24970.00	\$0.00
Choudhury , Ajit Engineering Howard University, Washington, DC	PhD 95-0840	RL/OC Detection Performance of Over Resolved Targets with Non-Uniform and Non-Gaussian	11/30/94 10/31/95	\$25000.00	\$0.00
Harackiewicz , Frances Electrical Engineering So. Illinois Univ-Carbondale, Carbondale, IL	PhD 95-0841	RL/ER Computer-Aided-Design Program for Solderless Coupling Between Microstrip and Str	01/01/95 12/31/95	\$23750.00	\$29372.00

1995 SREP SUB-CONTRACT DATA

Report Author Author's University	Author's Degree	Sponsoring Lab	Performance Period	Contract Amount	Univ. Cost Share
Losiewicz , Beth Psycholinguistics Colorado State University, Fort Collins, CO	PhD 95-0842	RL/IR Spanish	01/01/95 12/31/95 Dialect Identification Project	\$25000.00	\$4850.00
Musavi , Mohamad University of Maine, Orono, ME	PhD 95-0843	RL/IR Automatic Image Registration Using Digital Terrain Elevation Data	01/01/95 12/31/95	\$25000.00	\$12473.00
Norgard , John Elec & Comp Engineering Univ of Colorado-Colorado Sprg, Colorado	PhD 95-0844	RL/ER Infrared Images of Electromagnetic Fields	01/01/95 12/31/95	\$25000.00	\$2500.00
Richardson , Dean Photonics SUNY Institute of Technology, Utica, NY	PhD 95-0845	RL/OC Femtosecond Pump-Probe Spectroscopy System	01/01/95 12/31/95	\$25000.00	\$15000.00
Ryder, Jr. , Daniel Chemical Engineering Tufts University, Medford, MA	PhD 95-0846	RL/ER Synthesis and Properties B-Diketonate-Modified Heterobimetallic Alkoxides	01/01/95 12/31/95	\$25000.00	\$0.00
Zhang , Xi-Cheng Physics Rensselaer Polytechnic Institu, Troy, NY	PhD 95-0847	RL/ER Optoelectronic Study of Seniconductor Surfaces and Interfaces	01/01/95 12/31/95	\$25000.00	\$0.00
Drost-Hansen , Walter Chemistry University of Miami, Coral Gables, FL	PhD 95-0875	WHMC/ Biochemical & Cell Physiological Aspects of Hyperthermia	01/01/95 12/31/95	\$25000.00	\$8525.00
Baginski , Michael Electrical Engineering Auburn University, Auburn, AL	PhD 95-0869	WL/MN An Investigation of the Heating and Temperature Distribution in Electrically Exc	01/01/95 12/31/95	\$24995.00	\$10098.00
Berdichevsky , Victor Aerospace Engineering Wayne State University, Detroit, MI	PhD 95-0849	WL/FI Micromechanics of Creep in Metals and Ceramics at High Temperature	01/01/95 12/31/95	\$25000.00	\$0.00
Buckner , Steven Chemistry Colullmbus College, Columbus, GA	PhD 95-0850	WL/PO Development of a Fluorescenece-Based Chemical Sensor for Simultaneous Oxygen Qua	01/01/95 12/31/95	\$24900.00	\$8500.00
Carroll , James Electrical Engineering Clarkson University, Potsdam, NY	PhD 95-0881	WL/PO Development of High-Performance Active Dynamometer System for Machines and Drive	01/01/95 12/31/95	\$24944.00	\$38964.00
Choate , David Mathematics Transylvania University, Lexington, KY	PhD 95-0851	WL/AA SOLVING $z(t)=ln\{ /A[\cos(wlt)]+B[\sin(w2t)]+C\}$	01/01/95 12/31/95	\$24993.00	\$8637.00
Clarson , Stephen Materials Sci & Eng University of Cincinnati, Cincinnati, OH	PhD 95-0852	WL/ML Synthesis, Processing and Characterization of Nonlinear Optical Polymer Thin Fil	12/01/94 11/30/95	\$25000.00	\$15000.00
Cone , Milton Comp Science & Elec Eng Embry-Riddel Aeronautical Univ, Prescott, AZ	PhD 95-0853	WL/AA An Investigation of Planning and Scheduling Algorithms for Sensor Management	01/01/95 12/31/95	\$25000.00	\$11247.00
Courter , Robert Mechanical Engineering Louisiana State University, Baton Rouge, LA	PhD 95-0854	WL/MN A Study to Determine Wave Gun Firing Cycles for High Performance Model Launches	01/01/95 12/31/95	\$25000.00	\$3729.00

1995 SREP SUB-CONTRACT DATA

Report Author Author's University	Author's Degree	Sponsoring Lab	Performance Period		Contract Amount	Univ. Cost Share
Dominic , Vincent Electro Optics Program University of Dayton, Dayton, OH	PhD 95-0868	WL/ML	01/01/95	12/31/95	\$25000.00	\$12029.00
		Characterization of Electro-Optic Polymers				
Fadel , Georges Dept of Mechanical Engr Clemson University, Clemson, SC	PhD 95-0855	WL/MT	01/01/95	12/31/95	\$25000.00	\$8645.00
		A Methodology for Affordability in the Design Process				
Gould , Richard Mechanical Engineering North Carolina State Univ, Raleigh, NC	PhD 95-0856	WL/PO	01/01/95	12/31/95	\$24998.00	\$9783.00
		Data Reduction and Analysis for laser Doppler Velocimetry				
Hardie , Russell Electrical Engineering Univcity of Dayton, Dayton, OH	PhD 95-0882	WL/AA	01/01/95	12/31/95	\$24999.00	\$7415.00
		Hyperspectral Target Identification Using Bomen Spectrometer Data				
Hodel , Alan Electrical Engineering Auburn University, Auburn, AL	PhD 95-0870	WL/MN	01/01/95	12/31/95	\$24990.00	\$9291.00
		Robust Falut Tolerant Control: Fault Detection and Classification				
Janus , Jonathan Aerospace Engineering Mississippi State University, Mississippi State,	PhD 95-0871	WL/MN	01/01/95	12/31/95	\$25000.00	\$7143.00
		Multidimensional Algorithm Development & Analysis				
Jasiuk , Iwona Dept of Materials Science Michigan State University, East Lansing, MI	PhD 95-0857	WL/ML	01/01/95	12/31/95	\$25000.00	\$0.00
		Characterization of Interfaces in Metal-Matrix Composites				
Jouny , Ismail Electrical Engineering Lafayette College, Easton, PA	PhD 95-0880	WL/AA	01/01/95	12/31/95	\$24300.00	\$5200.00
		TSI Mitigation: A Mountaintop Database Study				
Li , Jian Electrical Engineering University of Florida, Gainesville, FL	PhD 95-0859	WL/AA	10/10/95	12/31/95	\$25000.00	\$4000.00
		Comparative Study and Performance Analysis of High Resolution SAR Imaging Techni				
Lin , Chun-Shin Electrical Engineering University of Missouri-Columbi, Columbia, MO	PhD 95-0883	WL/FI	01/01/95	12/31/95	\$25000.00	\$2057.00
		Prediction of Missile Trajectory				
Lin , Paul Mechanical Engineering Cleveland State University, Cleveland, OH	PhD 95-0860	WL/FI	01/01/95	12/31/95	\$25000.00	\$6886.00
		Three Dimensional Deformation Comparison Between Bias and Radial Aircraft Tires				
Liou , Juin Electrical Engineering University of Central Florida, Orlando, FL	PhD 95-0876	WL/EL	01/01/95	12/31/95	\$25000.00	\$11040.00
		Investigation of AlGaAs/GaAs Heterojunction Bipolar Transister Reliability Based				
Nandhakumar , Nagaraj Electrical Engineering University of Virginia, Charlottesville, VA	PhD 95-0861	WL/AA	01/01/95	12/31/95	\$24979.00	\$4500.00
		Thermophysical Invariants fro, LWIR Imagery for ATR				
Pasala , Krishna Dept of Electrical Engr University of Dayton, Dayton, OH	PhD 95-0879	WL/AA	01/01/95	12/31/95	\$25000.00	\$1078.00
		Effect of Electromagmetic Enviornment on Array Signal Processing				
Perkowski , Marek Dept of Electrical Engr Portland State University, Portland, OR	PhD 95-0878	WL/AA	01/01/95	09/15/95	\$24947.00	\$18319.00
		Functional Decomposition of Binary, Multiple-Valued, & Fuzzy Logic				

1995 SREP SUB-CONTRACT DATA

Report Author Author's University	Author's Degree	Sponsoring Lab	Performance Period		Contract Amount	Univ. Cost Share
Reeves , Stanley Dept of Electrical Engr Auburn University, Auburn, AL	PhD 95-0862	WL/MN	01/01/95	12/31/95	\$25000.00	\$0.00
			Superresolution of Passive Millimeter-Wave Imaging			
Rule , William Engineering Mechanics University of Alabama, Tuscaloosa, AL	PhD 95-0872	WL/MN	01/01/95	12/31/95	\$24968.00	\$14576.00
			Development of a Penetrator Optimizer			
Schauer , John Mech & Aerosp Eng University of Dayton, Dayton, OH	PhD 95-0873	WL/PO	11/01/94	11/30/95	\$25000.00	\$7428.00
			Heat Transfer for Turbine Blade Film Cooling with Free Stream Turbulence - Measu			
Schwartz , Carla Electrical Engineering University of Florida, Gainesville, FL	PhD 95-0863	WL/FI	01/01/95	12/31/95	\$25000.00	\$0.00
			Neural Network Identification and Control in Metal Forging			
Simon , Terrence Dept of Mechanical Engineering University of Minnesota, Minneapolis, MN	PhD 95-0864	WL/PO	01/01/95	12/31/95	\$24966.00	\$3996.00
			Documentation of Separating and Separated Boundary Layer Flow, for Application			
Skowronski , Marek Solid State Physics Carnegie Melon University, Pittsburgh, PA	PhD 95-0865	WL/EL	01/01/95	12/31/95	\$25000.00	\$6829.00
			Transmission Electron Microscopy of Semiconductor Heterojunctions			
Thirunarayan , Krishnaprasad Computer Science Wright State University, Dayton, OH	PhD 95-0866	WL/EL	01/01/95	12/31/95	\$25000.00	\$2816.00
			VHDL-93 Parser in SWI-PROLOG: A Basis for Design Query System			
Trelease , Robert Dept of Anatomy & Cell Bi University of California, Los Angeles, CA	PhD 95-0867	WL/ML	12/01/94	12/01/95	\$25000.00	\$0.00
			Development of Qualitative Process Control Discovery Systems for Polymar Composi			
Tsai , Chi-Tay Engineering Mechanics Florida Atlantic University, Boca Raton, FL	PhD 95-0874	WL/MN	01/01/95	12/31/95	\$24980.00	\$0.00
			Improved Algorithm Development of Massively Parallel Epic Hydrocode in Cray T3D			
Lewis , John Materials Science Engrng University of Kentucky, Lexington, KY	MS 95-0858	WL/MN	01/01/95	12/31/95	\$25000.00	\$13833.00
			The Characterization of the Mechanical Properties of Materials in a Biaxial Stre			

APPENDIX 1:
SAMPLE SREP SUBCONTRACT

**AIR FORCE OFFICE OF SCIENTIFIC RESEARCH
1995 SUMMER RESEARCH EXTENSION PROGRAM
SUBCONTRACT 95-0837**

BETWEEN

Research & Development Laboratories
5800 Uplander Way
Culver City, CA 90230-6608

AND

Florida Atlantic University
Department of Electrical Engineering
Boca Raton, FL 33431

REFERENCE: Summer Research Extension Program Proposal 95-0837
Start Date: 01-01-95 End Date: 12-31-95
Proposal Amount: \$25,000.00

- (1) **PRINCIPAL INVESTIGATOR:** Dr. Valentine A. Aalo
Department of Electrical Engineering
Florida Atlantic University
Boca Raton, FL 33431
- (2) **UNITED STATES AFOSR CONTRACT NUMBER:** F49620-93-C-0063
- (3) **CATALOG OF FEDERAL DOMESTIC ASSISTANCE NUMBER (CFDA):**12.800
PROJECT TITLE: AIR FORCE DEFENSE RESEARCH SOURCES PROGRAM
- (4) **ATTACHMENT 1 REPORT OF INVENTIONS AND SUBCONTRACT**
2 **CONTRACT CLAUSES**
3 **FINAL REPORT INSTRUCTIONS**

*****SIGN SREP SUBCONTRACT AND RETURN TO RDL*****

1. BACKGROUND: Research & Development Laboratories (RDL) is under contract (F49620-93-C-0063) to the United States Air Force to administer the Summer Research Program (SRP), sponsored by the Air Force Office of Scientific Research (AFOSR), Bolling Air Force Base, D.C. Under the SRP, a selected number of college faculty members and graduate students spend part of the summer conducting research in Air Force laboratories. After completion of the summer tour participants may submit, through their home institutions, proposals for follow-on research. The follow-on research is known as the Summer Research Extension Program (SREP). Approximately 61 SREP proposals annually will be selected by the Air Force for funding of up to \$25,000; shared funding by the academic institution is encouraged. SREP efforts selected for funding are administered by RDL through subcontracts with the institutions. This subcontract represents an agreement between RDL and the institution herein designated in Section 5 below.
2. RDL PAYMENTS: RDL will provide the following payments to SREP institutions:
 - 80 percent of the negotiated SREP dollar amount at the start of the SREP research period.
 - The remainder of the funds within 30 days after receipt at RDL of the acceptable written final report for the SREP research.
3. INSTITUTION'S RESPONSIBILITIES: As a subcontractor to RDL, the institution designated on the title page will:

- a. Assure that the research performed and the resources utilized adhere to those defined in the SREP proposal.
- b. Provide the level and amounts of institutional support specified in the SREP proposal..
- c. Notify RDL as soon as possible, but not later than 30 days, of any changes in 3a or 3b above, or any change to the assignment or amount of participation of the Principal Investigator designated on the title page.
- d. Assure that the research is completed and the final report is delivered to RDL not later than twelve months from the effective date of this subcontract, but no later than December 31, 1998. The effective date of the subcontract is one week after the date that the institution's contracting representative signs this subcontract, but no later than January 15, 1998.
- e. Assure that the final report is submitted in accordance with Attachment 3.
- f. Agree that any release of information relating to this subcontract (news releases, articles, manuscripts, brochures, advertisements, still and motion pictures, speeches, trade associations meetings, symposia, etc.) will include a statement that the project or effort depicted was or is sponsored by: Air Force Office of Scientific Research, Bolling AFB, D.C.
- g. Notify RDL of inventions or patents claimed as the result of this research as specified in Attachment 1.
- h. RDL is required by the prime contract to flow down patent rights and technical data requirements to this subcontract. Attachment 2 to this subcontract

contains a list of contract clauses incorporated by reference in the prime contract.

4. All notices to RDL shall be addressed to:

RDL AFOSR Program Office
5800 Uplander Way
Culver City, CA 90230-6609

5. By their signatures below, the parties agree to provisions of this subcontract.



Abe Sopher
RDL Contracts Manager

Signature of Institution Contracting Official

Typed/Printed Name

Date

Title

Institution

Date/Phone

ATTACHMENT 2
CONTRACT CLAUSES

This contract incorporates by reference the following clauses of the Federal Acquisition Regulations (FAR), with the same force and effect as if they were given in full text. Upon request, the Contracting Officer or RDL will make their full text available (FAR 52.252-2).

<u>FAR CLAUSES</u>	<u>TITLE AND DATE</u>
52.202-1	DEFINITIONS
52.203-3	GRATUITIES
52.203-5	COVENANT AGAINST CONTINGENT FEES
52.203-6	RESTRICTIONS ON SUBCONTRACTOR SALES TO THE GOVERNMENT
52.203-7	ANTI-KICKBACK PROCEDURES
52.203-8	CANCELLATION, RECISSION, AND RECOVERY OF FUNDS FOR ILLEGAL OR IMPROPER ACTIVITY
52.203-10	PRICE OR FEE ADJUSTMENT FOR ILLEGAL OR IMPROPER ACTIVITY
52.203-12	LIMITATION ON PAYMENTS TO INFLUENCE CERTAIN FEDERAL TRANSACTIONS
52.204-2	SECURITY REQUIREMENTS
52.209-6	PROTECTING THE GOVERNMENT'S INTEREST WHEN SUBCONTRACTING WITH CONTRACTORS DEBARRED, SUSPENDED, OR PROPOSED FOR DEBARMENT
52.212-8	DEFENSE PRIORITY AND ALLOCATION REQUIREMENTS
52.215-2	AUDIT AND RECORDS - NEGOTIATION
52.215-10	PRICE REDUCTION FOR DEFECTIVE COST OR PRICING DATA

52.215-12	SUBCONTRACTOR COST OR PRICING DATA
52.215-14	INTEGRITY OF UNIT PRICES
52.215-8	ORDER OF PRECEDENCE
52.215.18	REVERSION OR ADJUSTMENT OF PLANS FOR POSTRETIREMENT BENEFITS OTHER THAN PENSIONS
52.222-3	CONVICT LABOR
52.222-26	EQUAL OPPORTUNITY
52.222-35	AFFIRMATIVE ACTION FOR SPECIAL DISABLED AND VIETNAM ERA VETERANS
52.222-36	AFFIRMATIVE ACTION FOR HANDICAPPED WORKERS
52.222-37	EMPLOYMENT REPORTS ON SPECIAL DISABLED VETERAN AND VETERANS OF THE VIETNAM ERA
52.223-2	CLEAN AIR AND WATER
52.223-6	DRUG-FREE WORKPLACE
52.224-1	PRIVACY ACT NOTIFICATION
52.224-2	PRIVACY ACT
52.225-13	RESTRICTIONS ON CONTRACTING WITH SANCTIONED PERSONS
52.227-1	ALT. I - AUTHORIZATION AND CONSENT
52.227-2	NOTICE AND ASSISTANCE REGARDING PATIENT AND COPYRIGHT INFRINGEMENT

52.227-10	FILING OF PATENT APPLICATIONS - CLASSIFIED SUBJECT MATTER
52.227-11	PATENT RIGHTS - RETENTION BY THE CONTRACTOR (SHORT FORM)
52.228-7	INSURANCE - LIABILITY TO THIRD PERSONS
52.230-5	COST ACCOUNTING STANDARDS - EDUCATIONAL INSTRUCTIONS
52.232-23	ALT. I - ASSIGNMENT OF CLAIMS
52.233-1	DISPUTES
52.233-3	ALT. I - PROTEST AFTER AWARD
52.237-3	CONTINUITY OF SERVICES
52.246-25	LIMITATION OF LIABILITY - SERVICES
52.247-63	PREFERENCE FOR U.S. - FLAG AIR CARRIERS
52.249-5	TERMINATION FOR CONVENIENCE OF THE GOVERNMENT (EDUCATIONAL AND OTHER NONPROFIT INSTITUTIONS)
52.249-14	EXCUSABLE DELAYS
52.251-1	GOVERNMENT SUPPLY SOURCES

DOD FAR CLAUSES**DESCRIPTION**

252.203-7001	SPECIAL PROHIBITION ON EMPLOYMENT
252.215-7000	PRICING ADJUSTMENTS
252.233-7004	DRUG FREE WORKPLACE (APPLIES TO SUBCONTRACTS WHERE THERE IS ACCESS TO CLASSIFIED INFORMATION)
252.225-7001	BUY AMERICAN ACT AND BALANCE OF PAYMENTS PROGRAM
252.225-7002	QUALIFYING COUNTRY SOURCES AS SUBCONTRACTS
252.227-7013	RIGHTS IN TECHNICAL DATA - NONCOMMERCIAL ITEMS
252.227-7030	TECHNICAL DATA - WITHOLDING PAYMENT
252.227-7037	VALIDATION OF RESTRICTIVE MARKINGS ON TECHNICAL DATA
252.231-7000	SUPPLEMENTAL COST PRINCIPLES
252.232-7006	REDUCTIONS OR SUSPENSION OF CONTRACT PAYMENTS UPON FINDING OF FRAUD

APPENDIX 2:

SAMPLE TECHNICAL EVALUATION FORM

SUMMER RESEARCH EXTENSION PROGRAM TECHNICAL EVALUATION

SREP NO: 95-0811

SREP PRINCIPAL INVESTIGATOR: Dr. Michael Burke

Circle the rating level number, 1 (low) through 5 (high), you feel best evaluate each statement and return the completed form by mail to:

RDL
Attn: 1995 SREP Tech Evals
5800 Uplander Way
Culver City, CA 90230-6608
(310) 216-5940 or (800) 677-1363

- | | | |
|-----|---|-----------|
| 1. | This SREP report has a high level of technical merit. | 1 2 3 4 5 |
| 2. | The SREP program is important to accomplishing the lab's mission. | 1 2 3 4 5 |
| 3. | This SREP report accomplished what the associate's proposal promised. | 1 2 3 4 5 |
| 4. | This SREP report addresses area(s) important to the USAF. | 1 2 3 4 5 |
| 5. | The USAF should continue to pursue the research in this SREP report. | 1 2 3 4 5 |
| 6. | The USAF should maintain research relationships with this SREP associate. | 1 2 3 4 5 |
| 7. | The money spent on this SREP effort was well worth it. | 1 2 3 4 5 |
| 8. | This SREP report is well organized and well written. | 1 2 3 4 5 |
| 9. | I'll be eager to be a focal point for summer and SREP associates in the future. | 1 2 3 4 5 |
| 10. | The one-year period for complete SREP research is about right. | 1 2 3 4 5 |

11. If you could change any one thing about the SREP program, what would you change.

12. What would you definitely NOT change about the SREP program?

USE THE BACK FOR ANY ADDITIONAL COMMENTS.

Laboratory: Armstrong Laboratory
Lab Focal Point: Linda Sawin Office Symbol: AL/HRMI
Phone: (210) 536-3876

PERFORMANCE STUDY OF ATM-SATELLITE NETWORK

Dr. Valentine Aalo
Okechukwu Ugweje
Department of Electrical Engineering

Florida Atlantic University
Boca Raton, FL 33431

Final Report for:
Summer Research Extension Program
Rome Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, Washington, D.C.

and

Florida Atlantic University

December 1995

PERFORMANCE STUDY OF ATM-SATELLITE NETWORK

Valentine Aalo
Electrical Engineering Department
Florida Atlantic University

Okechukwu Ugweje
Electrical Engineering Department
Florida Atlantic University

Abstract

In order to investigate the effectiveness of using Asynchronous Transfer Mode (ATM) over satellite, we have carried out a study of satellite transmission of ATM cell streams. In this report, we describe our investigation of ATM over satellite with emphasis on the effect of the bit error characteristics of the satellite link on ATM cell streams. The effect of errors (both single bit and multiple bit bursty errors) have been characterized and studied. Exact and approximate expressions of the ATM quality of service (QoS) parameters such as cell loss ratio (CLR), cell error ratio (CER) and cell misinsertion rate (CMR) have been derived. Our study shows that the bit error probability of ATM transmission over satellite is affected more by bursty errors in the link than random single bit errors.

PERFORMANCE STUDY OF ATM-SATELLITE NETWORK

Valentine Aalo
Okechukwu Ugweje

1.0 INTRODUCTION

Asynchronous Transfer Mode (ATM) protocol is presently the focus of intense research interest. With its popularity still rising, ATM is poised to becoming the premier protocol for many communication and networking applications in the future. ATM was originally designed for transmission at high data rates on a physical medium with excellent error characteristics, such as optic-fiber links. However, it is possible to use ATM on a non-physical medium, such as the satellite link. The reliability afforded by optic-fiber links is not obtainable in satellite links. Many performance issues regarding the transfer of ATM information bearers via the satellite channel is still unresolved, especially at the Ka-band frequencies where transmission is severely impaired by atmospheric propagation effects, especially rain attenuation.

Currently, in most instances, separate networks are used to carry voice, data and video information mostly because these types of traffic have different transmission characteristics. For instance, data traffic tends to be bursty - not needing to communicate for an extended period of time and then suddenly needing to communicate large amount of information as fast as possible. On the other hand, voice and video tend to require even amount of information but are very sensitive to when and in what order the information is transmitted or received. Unlike currently used protocols, ATM promises the technical capabilities to handle any kind of information, voice, data, text, video and images, in an integrated manner. With ATM, separate network or protocol will not be necessary, and various speeds can be used for the various data streams.

The importance of satellites in modern day telecommunication is well known. Any meaningful communication at very long distances must involve a satellite. Satellites provide an essential part of global communication carrying large amount of information and offering several features not readily available with other means of communication. Satellites will play a vital role in the provision of ATM services. Many performance issues regarding the transfer of ATM information bearers via the satellite link remain unresolved. Thus, it is important to study the performance issues involved in using ATM over satellite communication links.

In this report, we present the result of the study of transmitting ATM via satellite. We adopt the following approach. Firstly, we present some background information on ATM-Satellite communication. This is a review

meant as an introduction to the concept of ATM over satellite. This review introduces ATM and satellite link independently and then discusses both subjects combined as an integral part of a communication system. The switching mechanism of an ATM switch is not discussed. The behavior of the satellite channel error characterization is then presented. Based on prior studies of satellite channels, the average probability of error in a nonlinear channel corrupted by multiple uplink and downlink interference, and noise is derived. Secondly, we present the problems encountered in using ATM protocols in a satellite channel. This is achieved by characterizing the ATM quality of service (QoS) parameters as functions of the satellite channel bit error probability. Only the major ATM QoS parameters such as cell error ratio (CER), cell loss ratio (CLR) and cell misinsertion rate (CMR) are analyzed. This research seeks to properly define these parameters in terms of the satellite channel bit error rate (BER). Finally, numerical results are presented from which some conclusions are drawn.

2.0 BACKGROUND INFORMATION

A. Asynchronous Transfer Mode (ATM)

ATM is a switching technology based on a cell-relay protocol that uses fixed-length packets as the basic transfer unit. The information to be transmitted is segmented into fixed length packets known as cells. Each cell has a fixed length of 53 bytes broken into two main sections - the header and the information field (also referred to as payload). The payload, consisting of 48 bytes, carries the actual information. The header consists of 5 bytes for addressing and error correcting mechanism. The cell header contains two transport mechanisms: the Virtual Path Identifier (VPI) and the Virtual Channel Identifier (VCI). Within this 5 bytes of header, 1 byte is used for control, 1 byte is used for header checksum error control and 3 bytes are used for the virtual identifier labels. The control field may also contain a bit to specify whether the cell is used for flow control or ordinary cell, an advisory bit to indicate whether this cell is given priority in case of traffic congestion, and so on. The payload may optionally contain 4 byte of ATM adaptation layer information and 44 bytes of actual data, or all 48 bytes may be information. The kind of payload is determined by a bit in the control field of the header. For more detailed specification of the ATM header and information field, see the appropriate standards committee's documents mentioned in [1].

In ATM protocol, information from one or several sources is multiplexed, buffered and formatted into cells as they are received and then transmitted asynchronously as soon as each cell is filled. The mode of information transfer is connection-oriented. This means that a virtual connection between the source and the destination user must be established before the information can progress. The fixed-length cell structure provides predictability in delay variance, so that applications requiring continuous bit-rate and variable bit-rate can be supported. The ATM cell structure is shown in Figure 1.

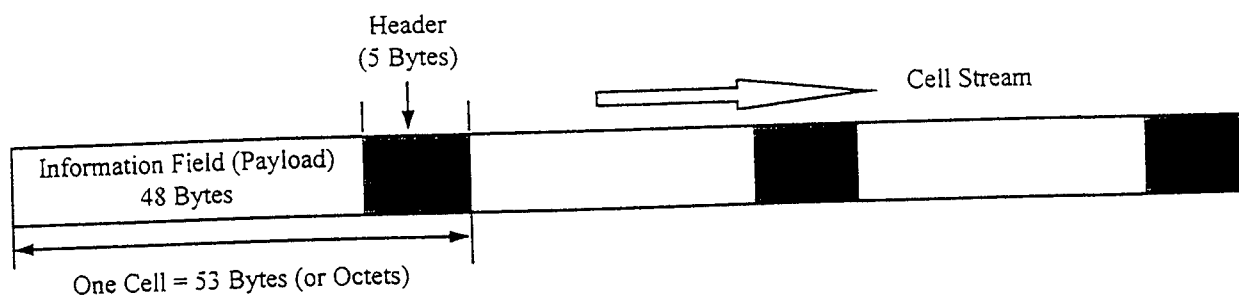


Figure 1. ATM Cell Structure

ATM is a layered structure allowing the possibility of multiplexing several services into a single network. To implement these features of ATM, three important lower level layers directly related to the characteristics of ATM protocol have been defined. These layers are the ATM Adaptation Layer (AAL), ATM Layer and the Physical Layer. These layers are briefly defined below.

ATM Adaptation Layer (AAL):

The use of ATM creates the need for an adaptation layer to support information transfer protocol of other services that are not based on ATM. The function of the AAL is to insert or extract information from non-ATM services and format all type of information signal into the 48 byte payload. This layer consist of two sublayers: the segmentation and reassembly (SAR) sublayer, and the convergence sublayer.

ATM Layer:

The ATM layer is concerned with transporting information across the network. This layer takes the information to be sent (received) and add (remove) the 5 byte header information ensuring that it is sent to (received from) the appropriate destination (source). It provides cell multiplexing, demultiplexing, and routing functions using the virtual part identifiers (VPI) and the virtual channel identifiers (VCI). Note that the virtual identifiers are simply the logical connections which determine the end-to-end connections.

Physical Layer:

This is the lowest layer of B-ISDN protocol stack which defines the electrical characteristics and network interfaces. It is at this layer where the ATM information is translated into electrical signals in the wire. The physical layer is actually independent of the protocol type, and can just as easily carry any type of signal as ATM cells.

B. Satellite Link

Because of the distances between a satellite and the earth stations, and given that the transmitted power diminishes as the square of the distance traveled, both uplink and downlink received signals are very weak, and can be easily disturbed by additive white Gaussian noise (AWGN) in the channel, signals transmitted from other earth stations, and signals from adjacent satellites. Spectrum congestion resulting from increased demand for telecommunications services have forced satellite designers to consider operating at very high frequencies. At these frequencies, atmospheric propagation effects are severe, particularly rain fades. It is commonly known that rain can severely attenuate satellite signals above 10 GHz. Furthermore, because of the power limitation of satellite systems, it is desirable to operate the transponder in saturation thereby introducing nonlinearity. The limited availability of satellite bandwidth implies that the transmitted signal must be severely bandlimited. This introduces intersymbol interference (ISI). These disturbances and nonlinear effects can significantly degrade satellite system performance.

The theoretical analysis of a satellite channel such as shown in Figure 2 could be found in several textbooks and will not be repeated here. However, the effect of the above nonlinearities and disturbances on the transmission of ATM through the satellite channel will be discussed. The most relevant performance measures in satellite transmission is the carrier-to-noise ratio (C/N) for analog communication and bit error probability for digital communication. If the channel is assumed to be analog, the noise power in a satellite communication link is determined not only by the noise power from the satellite itself, but also by the following factors:

1. Intermodulation in the satellite transponder $(\frac{C}{N})_{IM}$
2. Adjacent channel interference $(\frac{C}{N})_{IA}$
3. Cochannel interference $(\frac{C}{N})_{IC}$
4. Interference from other satellite systems $(\frac{C}{N})_{IO}$

Thus, the overall carrier-to-noise ratio of the satellite link can then be expressed as

$$\left(\frac{C}{N}\right)_T = \left[\left(\frac{C}{N}\right)_U^{-1} + \left(\frac{C}{N}\right)_D^{-1} + \left(\frac{C}{N}\right)_{IM}^{-1} + \left(\frac{C}{N}\right)_{IA}^{-1} + \left(\frac{C}{N}\right)_{IC}^{-1} + \left(\frac{C}{N}\right)_{IO}^{-1} \right]^{-1} \quad (1)$$

where the subscript U and D denotes the uplink and downlink respectively. In this analysis, the satellite channel is assumed to be a digital communication channel. Hence, the performance measure is stated in terms of the probability of bit error which will be calculated later.

C. ATM Over Satellite

In order to specify the transmission quality of ATM cells over satellite, the transmission performance will be evaluated by taking into consideration the channel interference, noise and nonlinearity effects. A model of an ATM-Satellite communication system is shown in Figure 2. The message block represents the source or sink of any kind of information such as voice, data, text, video or images. The ATM equipment converts the information to be sent into an ATM format ready to be transmitted through the satellite channel at the desired rate. The received signals from the satellite is reconstructed into an ATM format.

The internal operation of the ATM equipment is not emphasized in this study. It is assumed that the equipment generates or receives perfect ATM cells from any source. Furthermore, it is assumed that the ATM switching action supports a magnitude of interfaces required for satellite communication at different rates such as T1 (1.5 Mbps), T3 (45 Mbps), FDDI (100 Mbps) and SONNET OC-3 (155 Mbps). All the generating, analyzing, switching and multiplexing operation of ATM is modeled by this study. This assumption is in fact a realistic one because in Summer 1995, we conducted an experiment in Giffiss Air Force Base, Rome Laboratory, New York, using such an equipment manufactured by ADTECH, Inc.

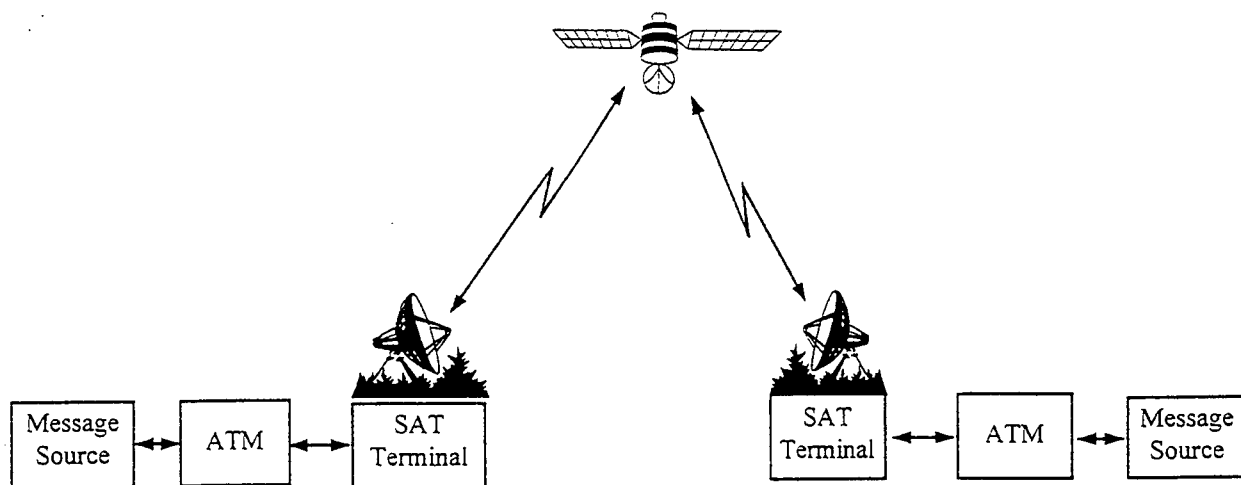


Figure 2. ATM-Satellite Communication System

The satellite transmission equipment consist of all the necessary satellite terrestrial equipment such as the baseband equipment, encoder/decoder, modem, up/down converters, HPAs and LPAs. These equipments are considered to be the standard satellite communication equipment with the added ability of interfaces that can receive and transmit ATM cells. This makes it possible to use already existing satellite equipments in the transmission of ATM information bearers.

A geostationary satellite channel may be modeled as a nonlinear AWGN channel. Such a channel is corrupted by multiple uplink and downlink interference and noise. Such a channel produces random single bit errors which depend on the received signal-to-noise ratio (SNR) or the energy to noise spectral density. Also, because of the forward error correction (FEC) codes used in satellite transmission, bursty errors are introduced in the channel. Since ATM was designed to be robust with single bit error, a thorough evaluation of the effect of bursty errors caused by the codes and channel nonlinearity on ATM signal transmission is important. To investigate this, the characteristic behavior of a nonlinear satellite channel is presented in the next section.

3.0 CHANNEL ERROR CHARACTERIZATION

The performance of the ATM network depends largely on the underlying physical layer which can be characterized among other things by the bit error rate (BER). As indicated earlier, ATM was designed for a physical medium with excellent BER characteristics, such as the fiber optic transmission link. It is assumed that errors in optic-fiber transmission systems are randomly distributed. However, in satellite communication, the forward error correction technique introduces burst errors. The length of the error bursts are randomly distributed. A number of features which were included in the older protocols to cope with unreliable channel are not part of ATM protocol. This implies less overhead requirements and increased throughput in optical networks. However, in error-prone channels such as the satellite link, this causes severe problems in reliable transmission of ATM cell bearers via satellite.

It is well known that in the presence of AWGN only, and assuming perfect phase coherence, when the SNR is relatively high, the system BER is well approximated by

$$P_e \approx 2Q\left(\sqrt{\frac{2E_s}{N_o}} \sin \frac{\pi}{M}\right) \quad (2)$$

where $E_s = (\log_2 M) E_b$ is the energy per symbol and $Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty \exp(-\frac{1}{2}\alpha^2) d\alpha$.

However, in addition to the AWGN, PSK signals that are transmitted through transponders of satellite systems must also contend with a number of impairments not normally encountered in terrestrial communication systems. These include cochannel and interchannel interference which are independent of the desired signal. Such interferences may arise from cross polarized signals from the same or adjacent satellites, sidelobe interference from ground stations transmitting to nearby satellites, intermodulation products, or from ground microwave link operating in the same frequency band as the received earth terminal. These interfering signals may corrupt the signal on both the uplink and downlink paths of the satellite channel. Furthermore, because of power limitations, transponders of the satellite system usually operate in saturation to ensure maximum radiated power. At saturation, the transponder operation is nonlinear and usually exhibits AM/AM and AM/PM conversions. Also, because of bandwidth limitations resulting from spectrum congestion, the transmitted signals are severely bandlimited leading to intersymbol interference (ISI). A combination of these effects lead to signal distortion and BER performance degradation.

The analysis of the MPSK performance in nonlinear satellite links have been done by a number of researchers [6], [7]. For analytical derivations, consider the ATM-Satellite communication model shown in Figure 3. The input and output characteristics of an ATM switch, including the internal operation, have been discussed in the literature. In [2], Wang and Frost presented the analysis of the cell blocking probability for an ATM switching system. Pattavina and Bruzzi [3], discussed the input and output behavior of a nonblocking ATM switch. The general properties, behavior, and quality of service parameters of the ATM protocol are presented in [4] and [5].

The operation of an ATM switch is not part of this study. We are only interested in the behavior of ATM cell streams in the satellite channel. It should be noted, however, that the input-output of an ATM switch is a sequence of binary bits. The output is a sequence of binary bit segmented into blocks of 53 bytes.

It is assumed that an arbitrary number of cochannel and interchannel interferers are present on the uplink and downlink. The following notations will be adopted:

N_u = Number of uplink cochannel interferers

$I_{u,i}$ = Power of uplink interferers ($i = 1, 2, \dots, N_u$)

N_d = Number of downlink cochannel interferers

$I_{d,i}$ = Power of downlink interferers ($i = 1, 2, \dots, N_d$)

$\psi_{d,i}$ = Phase of i -th downlink interferers ($i = 1, 2, \dots, N_d$)

σ_u^2 = Uplink AWGN power

σ_d^2 = Downlink AWGN power

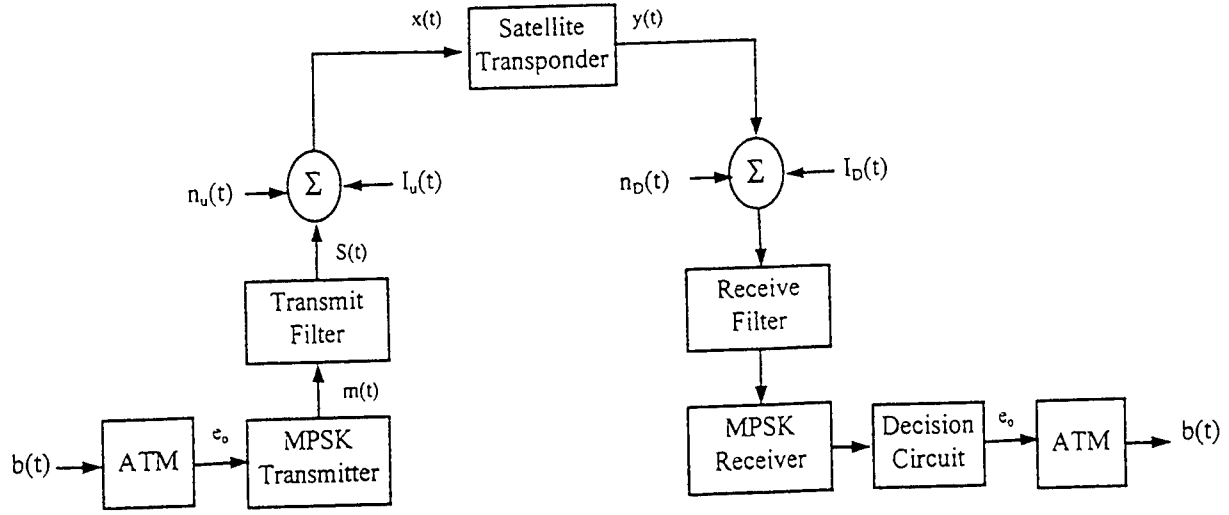


Figure 3. System Model for an MPSK ATM-Satellite Communication System

M-ary phase shift keying (MPSK) are usually considered in satellite communication systems because of its ability to contend the inherent power limitations of the satellite link. The MPSK transmitter modulates the block of binary bits into PSK signal which can be expressed as [6], [7]

$$m(t) = \sqrt{2A} \sum_{k=0}^{\infty} p(t - KT) \cos(w_c t + \theta_k) \quad (3)$$

where A is the transmitted signal power, T is the symbol duration time, w_c is the carrier frequency, θ_k is the transmitted phase taken from one of the M -phases, $\{\frac{i2\pi}{M}, i = 0, 1, 2, \dots, M-1\}$, and $p(t)$ is the unit pulse over one symbol time. This analysis will concentrate on two phases QPSK ($M=4$) and 8PSK ($M=8$). In Summer 1995, these modulation schemes were used in an ATM experiment over satellite. In this study, we compare the result of this study with the experimental results. The transmitted phases are assumed to have equal probability.

The transmitter is modeled by a time-invariant filter with impulse response

$$H(t) = 2h(t) \cos w_c t \quad (4)$$

After passing through the transmitter filter, the signal may be expressed as [7]

$$S(t) = \sqrt{2A} \left\{ q(t) \cos(w_c t + \theta_o) + \sum_{k=0}^{\infty} q(t - KT) \cos(w_c t + \theta_k) \right\} \quad (5)$$

where

$$q(t) = \int_0^{\infty} p(t - \tau) H(\tau) \exp(-jw_c \tau) d\tau$$

Along the channel, in the uplink direction, the transmitted signal is affected by interference $I_u(t)$ and noise $n_u(t)$. Thus, at the front end of the satellite transponder, the received signal can be expressed as the sum of the transmitted signal, noise and interference given by

$$x(t) = S(t) + I_u(t) + n_u(t) \quad (6)$$

where

$$\begin{aligned} I_u(t) &= \sum_{i=1}^{N_u} \sqrt{2I_{u,i}} \cos(w_c t + \psi_i(t)) \\ n_u(t) &= n_{uQ}(t) \cos w_c t - n_{uI}(t) \sin w_c t \end{aligned} \quad (7)$$

The noise $n_u(t)$ is a narrowband Gaussian noise with zero mean and variance $\sigma_u^2 = BN$, having the power spectral density $\frac{N}{2}$. For analytical purposes, (6) can be written as

$$x(t) = \sqrt{2AS(t)} \cos(w_c t + \theta_o + \theta(t)) + n_u(t) \quad (8)$$

where

$$\begin{aligned} S(t) &= q(t) \exp(j\theta) + \sum_{k \neq 0} q(t - KT) \exp[j(\theta_k - \theta_o - \theta)] \\ &\quad + \sum_{i=1}^{N_u} \sqrt{\left(\frac{I_{u,i}}{A}\right)} \exp[j(\psi_{u,i} - \theta_o - \theta)] \end{aligned}$$

Assuming that the front end filter at the satellite transponder is transparent, it can be shown that the input and output behavior of the transponder can be written as [7]

$$\begin{aligned} x(t) &= R(t) \cos(w_c t + \theta_o + \eta(t)) \\ y(t) &= f(R) \cos(w_c t + \theta_o + \eta(t)) + g(R) \end{aligned} \quad (9)$$

where R and η represent the complex envelope and phase of the narrowband Gaussian process and are both function of the uplink noise, ISI, and uplink and downlink cochannel interference. $f(R)$ and $g(R)$ represent the AM/AM conversions and AM/PM conversions, respectively. It is assumed that these two functions characterize the memoryless bandpass nonlinearity of the satellite transponder in which all harmonics that are generated are suppressed by appropriate filters. For example, Figure 4 shows the characteristics of a typical TWT in which the

AM/AM distortion has a saturation point at backoff power of 0 dB and the AM/PM distortion has a maximum slope at backoff power of -8 dB. The uplink and downlink characteristics are assumed to be identical. Thus on the downlink, the signal is also mixed with noise and interference.

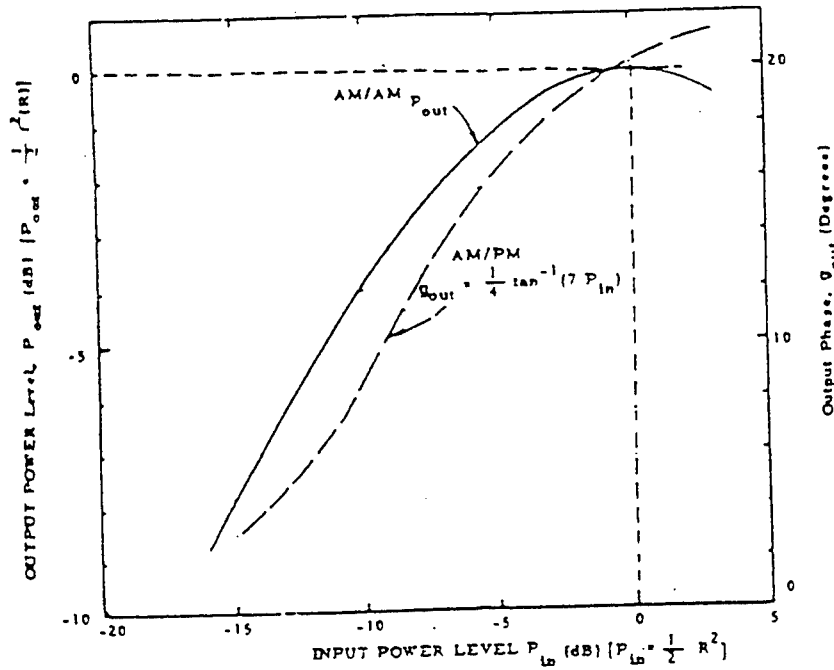


Figure 4. Transfer Characteristics of TWT

At the receiver, the MPSK signal is coherently demodulated by two local oscillators not shown. The decision circuit determines the transmitted phases based on the observation of the samples. Following the analysis in [6] and [7], and using a similar decision zone, the decision variables are a pair $\{r, \tilde{r}\}$ given as

$$\begin{aligned} r &= f(R_o) \cos(\theta_o + \eta_o + g(R_o)) + \sum_{i=1}^{N_d} \sqrt{2I_{d,i}} \cos \psi_{d,i} \\ \tilde{r} &= f(R_o) \sin(\theta_o + \eta_o + g(R_o)) + \sum_{i=1}^{N_d} \sqrt{2I_{d,i}} \sin \psi_{d,i} \end{aligned} \quad (10)$$

The receiver bases its decision on the pair $\{r, \tilde{r}\}$, where an error occurs if and only if the pair falls outside the decision zone of $\frac{2\pi}{M}$ radian centered at $\theta = 0^\circ$. The conditional bit error rate for MPSK, conditioned on the envelope R , the phase η , and the random phase of the downlink interferers, ψ_d , is given by [7]

$$P_e(R, \eta, \psi_d) = Q\left(\frac{d_{1M}}{\sqrt{2}\sigma_d}\right) + Q\left(\frac{d_{2M}}{\sqrt{2}\sigma_d}\right) - P_M(R, \eta, \psi_d) \quad (11)$$

where

$$\begin{aligned} d_{1M} &= f(R) \sin\left(\frac{\pi}{M} - \eta - g(R)\right) + \sum_{i=1}^{N_d} \sqrt{2I_{d,i}} \sin\left(\frac{\pi}{M} - \psi_{d,i}\right) \\ d_{2M} &= f(R) \sin\left(\frac{\pi}{M} + \eta + g(R)\right) + \sum_{i=1}^{N_d} \sqrt{2I_{d,i}} \sin\left(\frac{\pi}{M} + \psi_{d,i}\right) \\ d_{3M} &= \tan\left(\frac{2\pi}{M}\right) \left\{ f(R) \cos\left(\frac{\pi}{M} - \eta - g(R)\right) + \sum_{i=1}^{N_d} \sqrt{2I_{d,i}} \cos\left(\frac{\pi}{M} - \psi_{d,i}\right) \right\} \\ P_M(R, \eta, \psi_d) &= \frac{1}{\sqrt{2\pi}\sigma_d} \int_{d_{1M}}^{\infty} Q\left(\frac{\alpha - d_{1M} + d_{3M}}{\sigma_d \tan(\frac{2\pi}{M})}\right) \exp\left(-\frac{\alpha^2}{2\sigma_d^2}\right) d\alpha \end{aligned}$$

It should be pointed out that for large M and reasonably high SNR, $P_M(R, \eta, \psi_d)$ is usually very small compared with $P_e(R, \eta, \psi_d)$ and may sometimes be neglected. Finally, to obtain the average BER, $P_e(R, \eta, \psi_d)$ should be averaged over the statistics of R and η which, in turn, are functions of the uplink channel noise, ISI, and uplink and downlink cochannel interference. The expectation of $P_e(R, \eta, \psi_d)$ with respect to the independent random phase of the downlink cochannel interference may be evaluated by using a Taylor series expansion with a two-dimensional moment technique, such that the satellite bit error probability is given by [7]

$$p_e = E_{ISI, \psi_d} \{P_e(R, \eta, \psi_d)\} \quad (12)$$

4.0 ATM Quality of Service (QoS) PARAMETERS

In this section, the impact of transmission errors on ATM QoS parameters is presented. A number of ATM parameters have been identified as being very important in assessing the performance of the ATM protocol. The commonly used ones are the cell loss ratio (CLR), cell misinsertion rate (CMR) and cell errored ratio (CER). CLR is the ratio of the number of lost ATM cells sent by the user in a specific time interval to the total number of cells sent. Due to the multiplexing nature of the ATM cells, the limited size of the ATM buffer, the complex flow control

mechanisms, and of course the nature of the transmission channel, it happens that cells do get lost. CMR is the ratio of the cells delivered to the wrong destination to the total number of cells sent. It occurs as a result of an undetected error in the header that causes a change in the cell destination information in the header. CER errors occur in the payload of the ATM cells. Other parameters exist but those parameters may be considered as derivative of the above three QoS parameters. Each of these QoS parameters may be characterized in terms of the channel BER. The impact of channel errors on the ATM layer depends on the nature of the errors, whether they occur randomly or in bursts.

An ATM error characteristics behavior is shown in Figure 5. A header may have no errors, single bit error or multiple bit errors. ATM was designed to be robust with respect to random single bit errors. If there is single bit errors in the header, it is usually detected and corrected by the built in header error correction (HEC) mechanism. HEC can detect almost all types of errors (single bit and multiple bits), but can only correct single bits. Two events are possible when multiple bits in the header are in error. First, if the errors are detected, the entire cell is lost. This results in the ATM QoS parameter of cell loss ratio which compares the number of lost cells to the number of successfully received cells. On the other hand, if the errored bits are undetected, the cells are transmitted with false valid address. This leads to the misinsertion of cell into the wrong virtual address. The rate at which this misinsertion occurs is denoted by the cell misinsertion rate.

The payload has no error correcting mechanism. However, errors do occur in the payload. These errors cannot occur in the ATM layer. They occur in such layers as the ATM adaptation layer as illustrated in Figure 5. The errors that occur in the payload may be characterized by the QoS parameters such as the cell error ratio or the severely errored cell ratio (SECR).

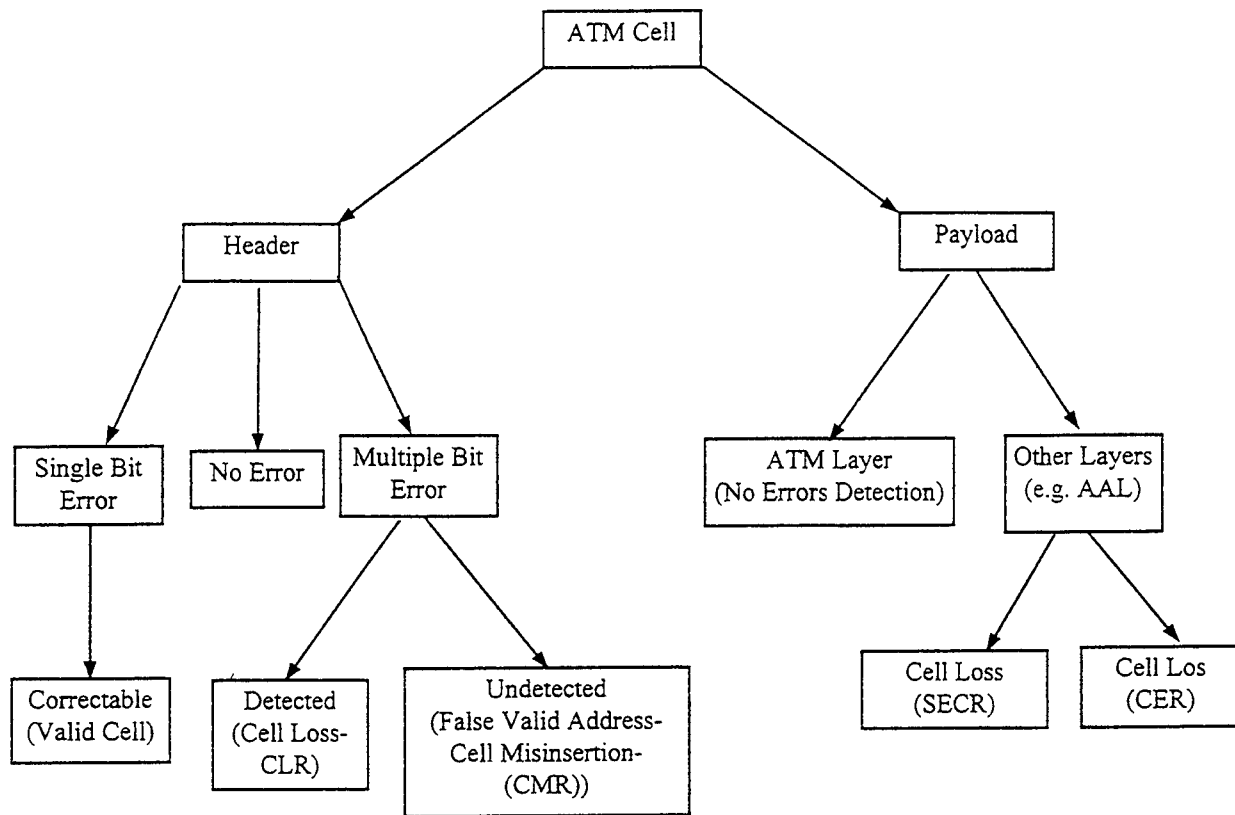


Figure 5. ATM Error Characterization

Generally, the error detection and correction mechanism is usually modeled by a Markov process. The simplest model is a two stage Markov process shown in Figure 6. The transition probabilities are not shown and are assumed to be equal from one stage to the other. At initialization, the receiver's error correcting mechanism is in default mode - correction mode. The receiver remains in this mode as long as no error is detected. If a single bit error is detected, the error is corrected and a transition is made to the detection mode. In this initialization mode, a multibit error will result in the entire cell being discarded. In the detection mode, no attempt is made to correct errors, and all errored cells are discarded. The ATM receiver remains in this mode as long as errored cells are received. When a header is examined and found to be error free, the receiver makes a transition to the correction mode.

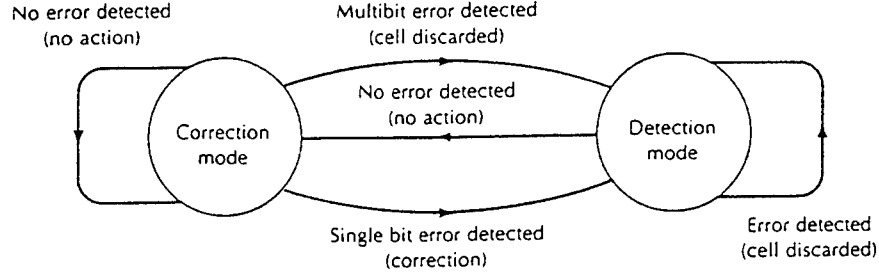


Figure 6. Operation Model of ATM Receiver

A. Random Single Bit Errors

The effect of single bit errors on the QoS parameters will now be examined. As each cell is received, the HEC calculation and comparison are performed. If a single bit error is detected, the receiver will correct it and makes a transition to the detection mode. If two consecutive cells are in error (single bit), the second cell is discarded because when the second error is detected, the receiver is already in the detection mode. We assume throughout that errors occur in the ATM-Satellite channel with BER of p_e . Note that this p_e has been computed in the previous section. Let k denote the number of bits in error and m the length of ATM header (i.e. $m = 5$ bytes (40 bits)). In this analysis, we assume that the HEC detects all header errors, both single and multiple bit errors. The probability that there are k bits in error in a header block of m bits is Binomially distributed and is given by [8]

$$P_{E_s}(k) = \binom{m}{k} p_e^k (1 - p_e)^{m-k} \quad (13)$$

It is evident that the probability for a single bit to be in error in a block of 40 bits is given by

$$P_{E_s} = 40 p_e (1 - p_e)^{39}$$

This is the probability of detecting single bit errors which is inherent in the ATM protocol.

Cell Loss Ratio (CLR):

A cell is lost when more than one bit is in error. CLR is the probability that more than one bit error occur in the ATM header (i.e. the probability that a cell is discarded). This probability may be expressed as

$$P_{CLR} = \sum_{k=2}^{40} \binom{40}{k} p_e^k (1 - p_e)^{40-k} \quad (14)$$

To account for burst errors using the dual mode Markov process, the detection and correction modes must be taken into consideration. For example, if the ATM receiver is in the correction mode and more than one error occurs, the second cell is discarded. Also, if the receiver is in the detection mode and at least one error occur, the cell is also discarded. Therefore, the cell loss probability is given by [8]

$$P_{CLR} = P_c \left[\sum_{k=2}^m \binom{m}{k} p_e^k (1-p_e)^{m-k} \right] + P_d [1 - (1-p_e)^m] \quad (15)$$

where $m=40$, is the number of bits in the header, and the subscripts c and d are used to denote the probabilities of the correction and detection modes respectively, given by

$$\begin{aligned} P_c &= (1-p_e)^m \\ P_d &= 1 - (1-p_e)^m \end{aligned}$$

Cell Error Ratio (CER):

CER is used to describe the errors that occur in the adaptation layer of the ATM protocol. It is the probability that at least one error occur in the payload. CER may be defined as

$$CER \equiv \frac{\text{Errored cell with at least one error in payload}}{\text{\# of successfully delivered cells}}$$

To compute this probability, the number of bits in the payload is used. Thus, the probability that one or more bits are in error in the payload is given by

$$P_{CER} = \Pr(\geq 1 \text{ error in payload}) = \sum_{k=1}^n \binom{n}{k} p_e^k (1-p_e)^{n-k} \quad (16)$$

where n is the total number of bits in the payload ($n = 384$). A cell is severely errored if the number of bits in error in the payload is much greater than one. Technically, the number of bits in error could vary between 2 to 384 for the severely errored cell ratio. If at least one bit of the payload is in error, the severely errored bits may be given by

$$P_{SECR} = \Pr(> \text{one error in payload}) = \sum_{k=2}^n \binom{n}{k} p_e^k (1-p_e)^{n-k} \quad (17)$$

where n is the total number of bits in the payload. In (16) and (17), we have neglected the fact that the ATM may use the dual mode operation in the header error control.

B. Burst Multibit Errors

Forward error correction (FEC) is usually employed on satellite communication in order to improve the power efficiency of the transmission system. Typically, convolutional codes are used interleaved with Viterbi

codes, leading to the occurrence of burst errors in the satellite link. Burst errors are also introduced by the FEC schemes adopted in the physical layer and the ATM layer. If an error burst affects more than one bit of the header, the error cannot be corrected. One of the subtle problems encountered in ATM transmission is that some of the multiple bit errors in the header may be improperly recognized as single bit error by the HEC resulting in improper correction. In this case, the cell is either discarded or transmitted with false valid address.

The distribution of the burst errors at the output of the decoder can be modeled in a number of ways. In one commonly used model, it is assumed that the error bursts as well as the errors in the burst are Poisson distributed. This leads to the Neyman-A Contagious model in which the probability that k errors occur in an interval of n bits is given by [7], [8]

$$P_{E_B}(k) = \frac{b^k}{k!} \exp\left(-\frac{np_e}{b}\right) \sum_{j=0}^{\infty} \left(\frac{np_e}{b} \exp(-b) \right)^j \frac{j^k}{j!} \quad (18)$$

where p_e is the channel bit error probability and b is the mean length of burst errored cell. For satellite communication, values of b range from 6 to 40.

Cell loss Ratio (CLR):

Similar to the single bit errors, CLR for burst errors is the probability that one bit error are affected by the burst error, which may be expressed as

$$\begin{aligned} P_{CLR} &= 1 - P_{E_B}(0) - P_{E_B}(1) \\ &= 1 - \beta \sum_{j=0}^{\infty} \left(\frac{mp_e}{b} \exp(-b) \right)^j \left[\frac{j+1}{j!} \right] \end{aligned} \quad (19)$$

where

$$\beta = \exp\left(-\frac{mp_e}{b}\right).$$

Cell Error Ratio (CER):

For CER, the probability that one or more errors occur in the payload is given by

$$\begin{aligned} P_{CER} &= 1 - \Pr(\text{no error in payload}) \\ &= 1 - \exp\left[-\frac{np_e}{b}(1 - \exp(-b))\right] \end{aligned} \quad (20)$$

Recall that n is the number of bits in the payload ($n = 384$), p_e is the channel bit error probability and b is the mean length of burst error.

Cell Misinsertion Rate(CMR):

An ATM header consist of 32 bits of information and 8 bits of HEC, a total of 40 bits. Thus, there are 2^{40} (*> one Trillion*) possible headers, out of which only 2^{32} (*> 4295 Million*) are valid headers. If a header is hit by an error burst, multiple bit errors will occur such that some of the headers will be invalid codeword. The probability that a valid codeword in the header is generated as the ratio of the valid codewords to the total codewords ($2^{32}/2^{40} = \frac{1}{256}$). Recall also that all single bit errors are correctable. This implies that 40 bit patterns of the valid codeword are correctable. Therefore, the probability that a valid correctable header is generated by the random bit pattern is the probability of the codeword multiplied by 41 (one correct codeword and 40 correctable codewords), is given by $41/256$. This means that a random bit pattern is considered a valid ATM header with the probability $41/256$.

5.0 NUMERICAL RESULTS

In this section, the numerical results are presented. At the data rate of 45 Mbps, the ATM QoS parameters such as the cell loss ratio (CLR), cell error ratio (CER) and cell misinsertion rate (CMR) are plotted against the satellite link bit error rate. In the computation of the bit error rate, two types of modulation were used, namely, QPSK and 8PSK. These plots are presented in Figures 7 through 9. It is evident from these graphs that generally, as the link BER increases, the performance of the ATM-Satellite system deteriorates. Figure 7 shows the QoS parameters as a function of the satellite channel BER for random single bit errors. All the parameters show degradation in performance as the BER increases. The behavior of the cell misinsertion rate shows some sign of instability. This is expected since the misinsertion of cells in false addresses is caused by many factors. In Figures 7 and 8, the performances are given for QPSK while in Figure 9, the results are for 8PSK.

In general, the ATM/Satellite system performance can further be enhanced by using an additional channel coding scheme to improve the satellite bit error performance. Specifically, Reed-Solomon coding with Viterbi decoding may be used. However, care must be taken since in this case, the concatenation of the forward error correction code (FEC, inner code) with the header error correction code (HEC, outer code) may result in a poor performance of the HEC code. Since, then, the outer code is only capable of correcting single bit errors, the errors at the output of the inner decoder must be dispersed by interleaving. To obtain best results, the ATM-cell headers may be interleaved for several cells so that the performance of the ATM over random bit error channel may be achieved. It should be pointed out that the full potentials of using additional coding on the satellite channel may be exploited further by replacing the ATM HEC code with a more powerful code, thereby improving the robustness to channel burst errors. We conclude, therefore, that the use of additional satellite channel coding (e.g. Reed-Solomon coding) either for

further protection of the ATM cell headers or to protect against single ATM cell errors will dramatically improve ATM QoS parameters such as cell loss rate.

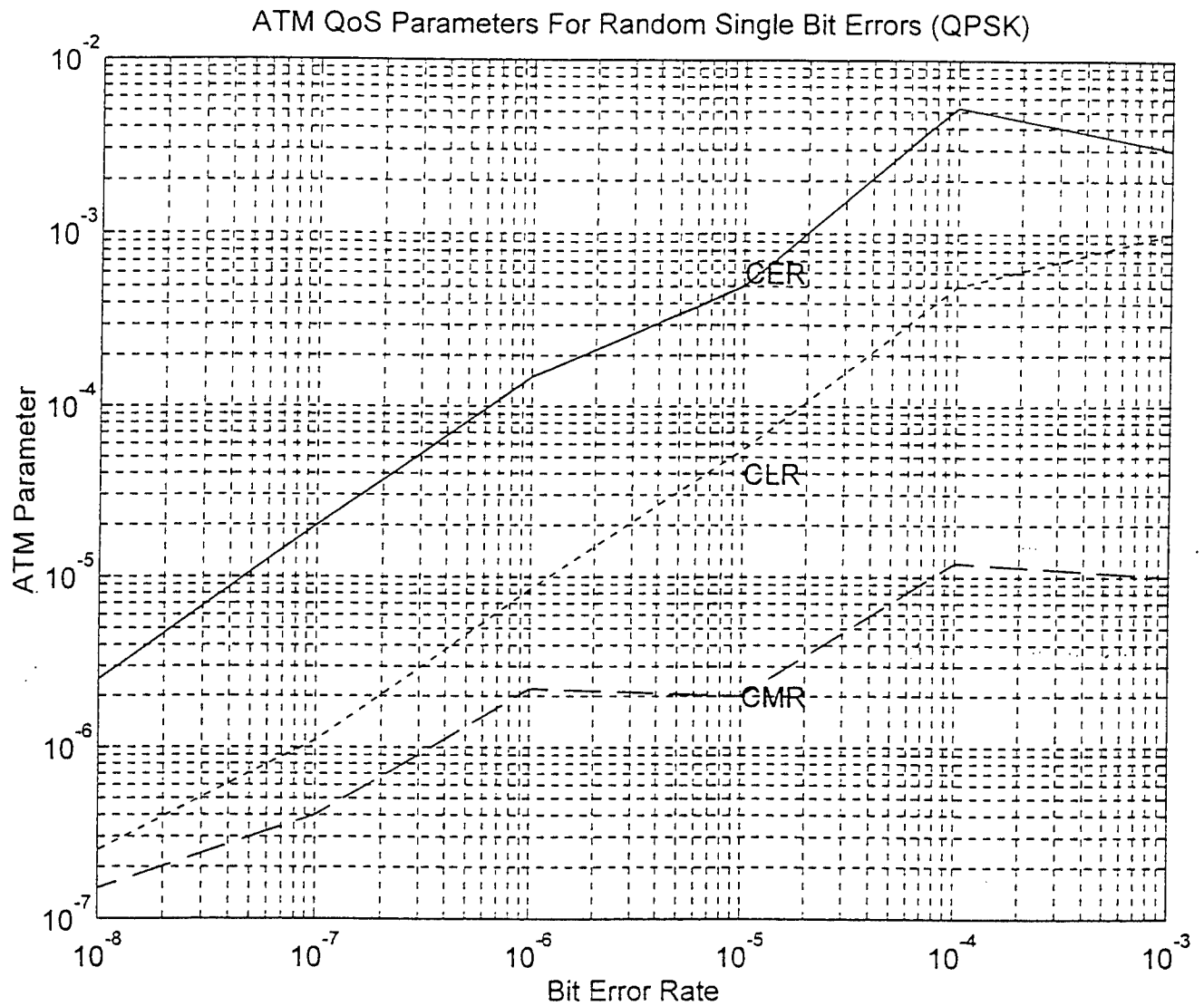


Figure 7: ATM QoS Parameters for Random Single Bit Errors (QPSK)

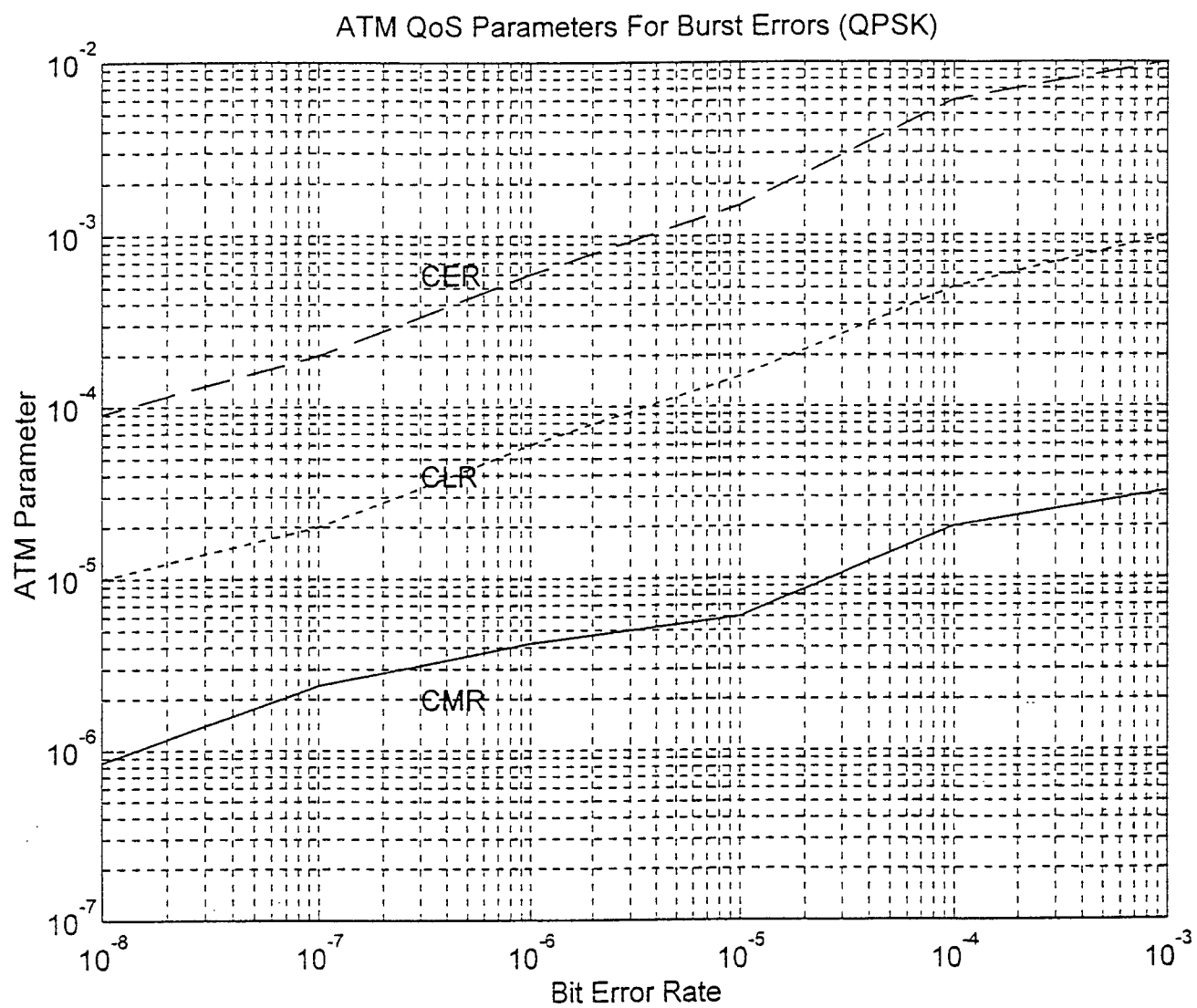


Figure 8: ATM QoS Parameters for Burst Errors (QPSK)

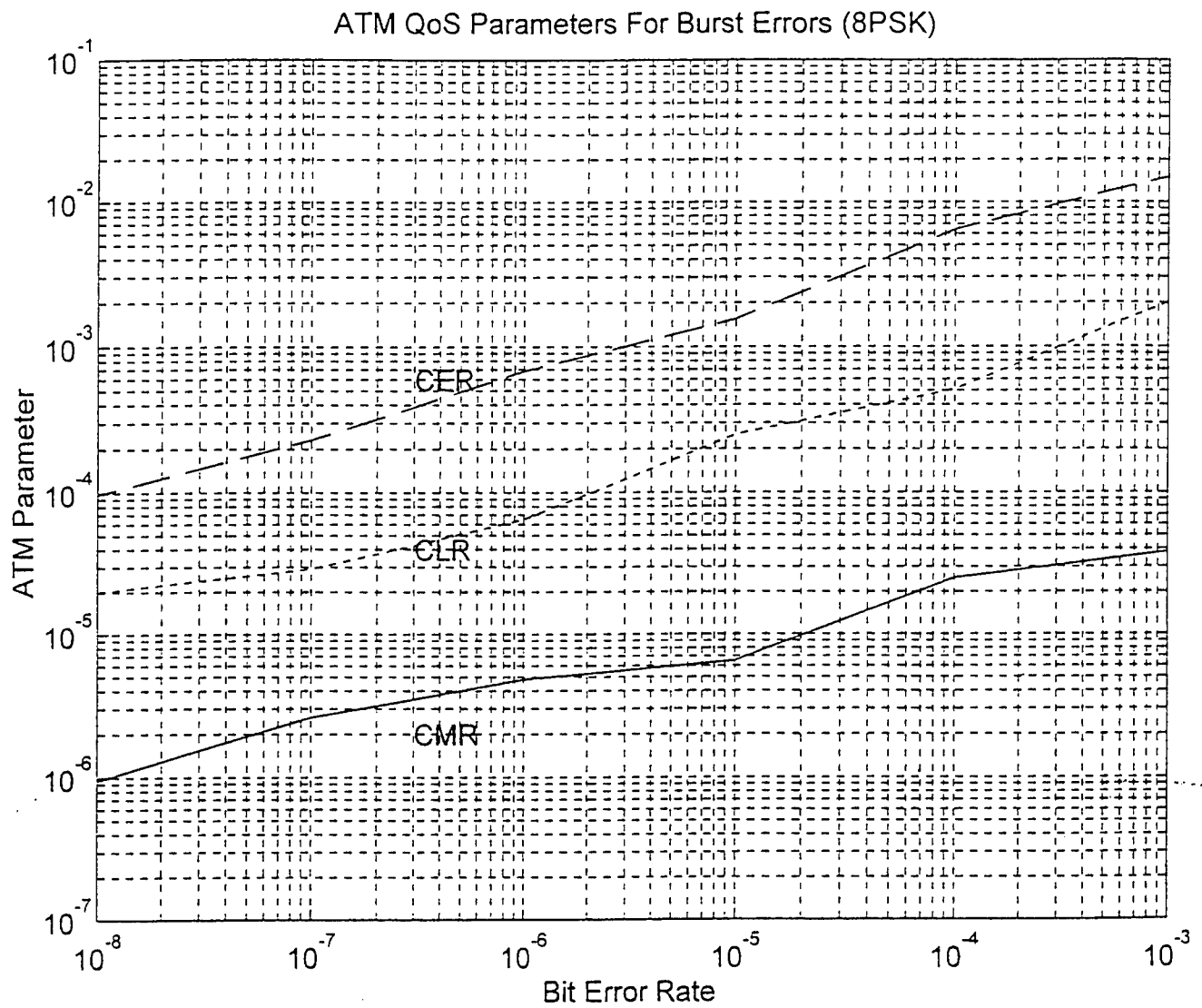


Figure 9: ATM QoS Parameters for Burst Errors (8PSK)

6.0 CONCLUSION

The use of the satellite transmission for ATM cells was investigated. The impact of single bit errors and bursty multiple bit errors on some ATM quality of service parameters, namely CLR, CMR, and CER was presented. First the transmission properties of the ATM protocol and the satellite link was discussed. The bit error probability for a nonlinear satellite link corrupted by multiple interference and additive white Gaussian noise was derived. Based on the satellite link BER, the probabilities of the ATM QoS parameters were derived.

REFERENCES

- [1] L. G. Guthbert and J-C Sapanel, *ATM The Broadband Telecommunications Solution*, England, Short Run Press Ltd., 1993.
- [2] Q. Wang and V. S. Frost, "Efficient estimation of cell blocking probability for ATM systems," *IEEE Trans. Networks*, vol. 1, No. 2, pp. 230-235, April 1993.
- [3] A. Pattavina and G. Bruzzi, "Analysis of input and output queuing for nonblocking ATM switches," *IEEE Trans. Networks*, vol. 1, No. 3, pp. 314-327, June 1993.
- [4] R.O. Onvural, *Asynchronous Transfer Mode Networks Performance Issues*, Boston; Artech House, 1994.
- [5] W. Stallings, *ISDN and Broadband ISDN with Frame Relay and ATM*, Third Edition, Prentice-Hall, Inc., Englewood Cliff, New Jersey, 1995.
- [6] T-C. Huang, J. K. Omura and W. C. Lindsey, "Analysis of coherent satellite communication system in the presence of interference and noise," *IEEE Trans. Commun.*, vol. COM-29, No. 5, pp. 593-604, May 1981.
- [7] I-K. Hwang and L. Kurz, "Digital data transmission over nonlinear satellite channels," *IEEE Trans. Commun.*, vol. COM-41, No. 11, pp. 1694-1702, November 1993.
- [8] S. Ranseier and T. Kaltenschnee, "ATM over satellite: Analysis of ATM QoS parameter," pp. 1562-1566, ICC 95.

INTERFERENCE EXCISION IN SPREAD SPECTRUM COMMUNICATION SYSTEMS
USING TIME-FREQUENCY DISTRIBUTIONS

Dr. Moeness G. Amin
Department of Electrical and Computer Engineering

Villanova University
Villanova, PA 19085

Final Report for:
Summer Research Extension Program
Rome Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, Washington, D.C.

and

Villanova University

December 1995

INTERFERENCE EXCISION IN SPREAD SPECTRUM COMMUNICATION SYSTEMS USING TIME-FREQUENCY DISTRIBUTIONS

Moeness G. Amin
Professor
Department of Electrical and Computer Engineering
Villanova University

Abstract

The capability of the time-frequency distributions (TFDs) to properly represent a single as well as multiple component signals in time and frequency permits the application of a new approach for interference excision in spread spectrum communication systems. In this report, the instantaneous frequency estimate from the TFD is used to construct a finite impulse response filter which, when applied to constant envelop nonstationary interference characterized by its instantaneous frequency, will substantially reduce its power with a minimum possible distortion of the desired signal. The proposed technique is therefore a case of open loop adaptive filtering. Three- and five- coefficient zero phase excision filters are considered. Closed form expressions of the improvement of SNR at the receiver correlator output using the TFD-based adaptive filtering are derived for two extreme cases of time-varying interference, namely those of fixed frequency sinusoids and randomly changing instantaneous frequency. Simulations results including the bit error rates are presented for both swept and frequency hopping jammers.

INTERFERENCE EXCISION IN SPREAD SPECTRUM COMMUNICATION SYSTEMS USING TIME-FREQUENCY DISTRIBUTIONS

Moeness G. Amin

Introduction

Spread spectrum (SS) systems are widely used in communications in a variety of applications including suppression of a strong interfering signal due to jamming or multipath propagation and in low probability of intercept communications. The spread spectrum system is characterized by 1) The signal occupies a bandwidth much in excess of the minimum bandwidth necessary to send the information. 2) Spreading is accomplished by means of spreading signal, often called a code signal, which is independent of the data. 3) At the receiver, despreading for recovering the original data is accomplished by the correlation of the received spread signal with a synchronized replica of the spread signal used to spread the information [1].

The most commonly used type of SS is the direct sequence (DS) in which modulation is achieved by superimposing a pseudorandom (PN) sequence $p(t)$ upon the data bits $d(t)$, as shown in Fig.1 below. The functions $n(t)$ and $i(t)$ are the channel noise and interference signals, respectively. T (T_c) is the width of information bit (PN chip). At the receiver, the cross correlation with the replica of the PN sequence transfers the information signal back to its original bandwidth, while reducing the level of a narrowband interference by spreading it across the bandwidth occupied by the PN sequence (see Fig.2).

The performance of a PN SS system can be further improved, with respect to its immunity to narrowband interference, by estimating the interference and subtracting it from the input data, as shown in Fig.3. This is commonly achieved by applying an excision filter prior to despreading (see Fig.4). This filter suppresses the interference and thus increases the signal to noise ratio at the output of the correlator. The filter coefficients can be provided using spectral estimation methods or adaptive filtering techniques. Although Fig. 4 shows the

excision filter implementations in the time-domain, interference excision can be performed in various domains, including[2]:

- 1) *Frequency Domain*: The FFT of the data over one information bit is weighted by appropriate values and then transformed back to the time domain[3,4]. This is an effective method for stationary narrow-band interference. Sidelobes may present a problem in removing the interference without losing some of the signal energy.
- 2) *Time domain*: This includes clipping or gating the high energy regions. It also includes Wiener Filtering, adaptive linear predictors and smoothers [3,5]. Tracking is highly dependent on the SNR and often fails under rapidly time-varying interference.
- 3) *Time and frequency domains*: A transversal filter is designed from the spectral information of the data[3]. Spectral estimation methods combined with open loop adaptive filtering have been shown to suffer from the same drawbacks as frequency-domain techniques.
- 4) *Wavelet /Gabor domain*: The discrete wavelet transform (DWT) or the Gabor transform is applied to the data and the coefficients of high energy are removed prior to the inverse transform[6]. The DWT is appropriate for cases of pulse jamming or interference with burst characteristics. The Gabor transform is an effective excision tool, only when the interference is consistent with the transform's corresponding tiling of the time-frequency plane[7]. The same is true for the wavelet transform.

None of the above methods is capable of properly incorporating the time-varying nature of the interference frequency characteristics. For illustrations, Fig.5 shows that the frequency- and time-domain excisions, in essence, respectively remove all desired signal information over the frequency band ΔF and time duration ΔT . As such, we maintain that in the case of a chirp as well as other time-varying interfering signals, frequency-domain methods ignore the fact that only few frequency bins are contaminated by the jammer at a given time. Time domain excision techniques on the other hand do not account for the interference characteristics where only few

time samples may be contaminated by the jammer for a given frequency. Applying either method will eliminate the interference, but unnecessarily reduce the desired signal energy.

The principle task is then to highly attenuate the received signal in those *time-frequency regions* which contain strong interference, as depicted by the region in between the fine lines in Fig.5. In this paper, we propose to employ time-frequency (t-f) distributions for performing time-varying spectral analysis on the received signal. We focus on the class of constant envelop nonstationary interference which is characterized by its instantaneous frequency (IF). This class includes swept (chirp) and frequency hopping jammers. On the basis of the instantaneous frequency estimate, two approaches, in essence, can be considered[2]. In the first approach, a linear phase open loop adaptive filter is applied to annihilate the time-varying interference with minimum loss of signal energy (see Fig.6-a). In the second approach, the interference is synthesized in the time-frequency plane using a mask centered around its instantaneous frequency. This result is fed to the received PN correlator(see Fig.6-b). Only, the first approach is considered in this report.

Time-Frequency Distributions: An Overview

Time-frequency distributions (TFDs) are being recognized as powerful tools in signal analysis and processing. Although time-frequency analysis has been of interest for some-time, several advances in recent years have broaden its application areas to include both civilian and military use. Over the past few years, research efforts have allowed proper depiction of spectral characteristics of signals with time-varying frequency contents. Separation between signals overlapping in time and in frequency, which could not be achieved using windowing or filtering techniques has thus become feasible using a large class of joint time-frequency distributions. Classification, detection, and estimation of signals can now be effectively performed in the time-frequency (t-f) plane and yields significant improvements over existing methods which are based on time domain or frequency domain processing.

TFDs are uniquely characterized by a two dimensional function, which is referred to as a "kernel". The t-f kernel can be designed such that the corresponding TFD satisfies several desired properties. For a full discussion of the time-frequency distributions and kernel design methods, we refer the reader to reference [8]. Among the desired t-f properties is the capability to satisfy the instantaneous frequency condition. Generally, this property allows the TFD to encounter peaks at the derivative of the phase of each signal component, irrespective of their time-varying nature.

The time-frequency distribution C_f of the signal $f(t)$ is defined as

$$C_f(t, \omega; \varphi) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \varphi(t-u, \tau) f(u+\tau/2) f^*(u-\tau/2) e^{-j\omega\tau} du d\tau$$

where "t" is the time index and "f" is the frequency index. The t-f kernel $\varphi(t, \tau)$ is a function of the time and lag variables. The well known Wigner distribution is a special case of the above equation with $\varphi(t, \tau) = \delta(\tau)$. A closer look at the above equation reveals the simple fact that the TFD is the Fourier transform (FT) of an estimated autocorrelation function. However, contrary to the common way of performing time-averaging, the dependency of $\varphi(t, \tau)$ on τ allows the autocorrelation function estimation to be different for different lags.

Fig.7 (a,b) shows the instantaneous frequency of a chirp and a frequency hopping signal along with the estimates obtained by using the Wigner distribution and the Born-Jordan distribution[8], respectively. It is clear how close the frequency estimates from the TFDs are to the exact values.

In addition to the instantaneous frequency, there are other common desired properties which qualify a TFD for proper representations of signals in time and frequency. These properties include the time support and frequency support. Both properties are important for the cases of pulse and bandlimited signals, since they, respectively, allow the TFD to be zero (shows no power) at all time instants and frequency bins where the signal is not present. The TFD should also satisfy the marginals properties in which the distribution of signal power over only the time variable or the frequency variable can be separately obtained from the joint TFD.

Each time-frequency property translates into a requirement (constraint) on the employed kernel. Recently, much attention is paid to the reduced interference distribution (RID) constraint where the two-dimensional Fourier transform of kernel $\varphi(t, \tau)$ is required to have low-pass filter characteristics. This condition is important to reduce the signal cross-terms, which arise due to the interaction of the signal components via the bilinear data products $f(t + \tau/2)f^*(t - \tau/2)$. These terms are not part of the original data. They are viewed as undesired components, since they obscure the signal autoterms (true components) and provide false interpretation of the distribution of the signal power over time and frequency. It is noted that the Wigner distribution kernel does not possess the RID property and thus produce a distribution with heavy cross-terms presence. All t-f constraints, including the RID, are compatible and can be satisfied simultaneously. Full discussions of the merits of the t-f properties and their significance to localize the signal in time and frequency can be found in [8]. Fast algorithms for on-line implementations of the TFDs are based on fast Fourier transform, recursive kernels, and singular value decomposition (SVD)[8].

The time-frequency distributions (TFDs) have been successfully applied in Biomedicine for studying cardiac sounds and EEG signals in epilepsy. In biological application, TFDs have been used to give more insights into marine mammal sounds, particularly concerning the fine structure. They have lead to increased understanding of the physiological anatomical and behavioral aspects on marine mammal communication. TFDs have also been used on a number of automotive signals, including engine sounds, door slam sounds, and wiper sounds to give better characterization of the unpleasant and annoying sound components. In industrial applications, TFDs have been applied to welding, bearing, tool cutting, and CMM signals. This paper is concerned with the applications of TFDs to communication signals.

Analysis

Based on the instantaneous frequency, two constraints should exist to construct an excision filter with desirable characteristics. First, an FIR filter with short impulse response must be used. Long extent filters are likely to span segments of changing frequency contents, and as such, allow some of the jammer components to escape to the filter output. Second, the filter frequency response must be close to an ideal notch filter to be able to null the interference with minimum possible distortion of the signal. This property, however, requires filters with infinite, or relatively long impulse response.

The above two conflicting requirements are present in most excision techniques dealing with time-varying signals [2]. In this paper, we work towards satisfying the first constraint and show that short excision filters, operating on the IF estimates from the TFDs, can be effectively used to improve the overall receiver performance under a large class of nonstationary interference.

The shortest excision filter is a three coefficient zero-phase filter which dedicates its complex conjugate zeros to notch the instantaneous frequency. Since there are numerous nonstationary jamming environments for which the effectiveness of this filter can be evaluated, we present below the receiver performance under the two extreme conditions: fixed and randomly changing instantaneous frequency conditions. In both cases, we highlight the main downfalls resulting from using the three coefficient filter and argue for two alternatives, namely the use of a five coefficient filter and two identical three coefficient filters. In the latter, one filter processes the received data, whereas the second operates on the PN at the receiver.

In the transmitted signal generation, the number of PN chips per information bit is L , i.e.,

$$b_k(t) = \sum_{j=1}^L p_{kj} q(t - jT_c) \quad (1)$$

where p_{kj} represent the output sequence from the PN code generator for the k th information bit and $q(t)$ is a rectangular pulse of duration T_c and unit energy. The total transmitted signal may be expressed in the form

$$s(t) = \sum_k I_k b_k(t - kT_b) \quad (2)$$

where $\{I_k\}$ represents the binary information sequence and $T_b = T = LT_c$ is the bit interval (reciprocal of the bit rate). The received signal has the form $r(t) = s(t) + i(t) + w(t)$, where $w(t)$ is the uncorrelated white noise process, $s(t)$ is the desired signal, and $i(t)$ the interference.

Case I: Three Coefficient Excision Filter

Let $h(n)$ represent the impulse response of the linear phase excision filter with coefficients $h(0)=h(1)=1$, $h(2)=-2 \cos(\omega_n)$, where ω_n is the instantaneous frequency. The filter zeros lie on the unit circle at $\exp(\pm j\omega_n)$. The filter input is the sampled signal $r(n)$ and the output is $y(n)$

$$y(n) = \sum_{m=-1}^1 h(m)r(n-m) \quad (3)$$

By linearity, the output can be written as

$$y(n) = p_o(n) + i_o(n) + w_o(n) \quad (4)$$

where the subscript "o" denotes output. The sequences $p(n)$, $i(n)$, and $w(n)$ are the sampled versions of the PN, interference, and the noise components of the input data over a given bit period. The output signal is fed into the PN correlator yielding the decision variable U ,

$$U = \sum_{n=1}^L p_o(n)p(n) + \sum_{n=1}^L i_o(n)p(n) + \sum_{n=1}^L w_o(n)p(n) \quad (5)$$

If it is assumed that the interference is entirely eliminated by filtering, then

$$U = \sum_{n=1}^L p_o(n)p(n) + \sum_{n=1}^L w_o(n)p(n) = U_1 + U_2$$

The first term U_1 is due to the desired signal and can be detailed as

$$\begin{aligned}
U_1 &= \sum_{n=1}^L \sum_{m=-1}^1 h(m) p(n-m) = \sum_{n=1}^L (p(n-1) + p(n+1) - 2 \cos(\omega_n) p(n)) p(n) \\
&= \sum_{n=1}^L p(n) p(n-1) + \sum_{n=1}^L p(n) p(n+1) - 2 \sum_{n=1}^L \cos(\omega_n) p^2(n)
\end{aligned} \tag{6}$$

The PN sequence is assumed to be identically distributed random variables such that $p(n)=1,-1$ with equal probability.

A. Random Changing IF

A highly nonstationary jamming environment is simulated by choosing a constant envelop interference whose instantaneous radian frequency is a random variable uniformly distributed over $[0, 2\pi]$. With this assumption, the expected value and the mean square value of U can be determined as follows

$$E[U_1] = \sum_{n=1}^L E[p(n)p(n-1)] + \sum_{n=1}^L E[p(n)p(n+1)] - 2 \sum_{n=1}^L E[\cos(\omega_n)] \tag{7}$$

But since $E[\cos(\omega_n)] = \frac{1}{2\pi} \int_0^{2\pi} \cos(\omega_n) d\omega_n = 0,$

then $E[U_1] = 0.$ (8)

Further, it can easily be shown that $E[U_2] = 0$. The SNR of the correlator is defined as the ratio of the square of the mean to the variance of the correlator output y_{peak} (peak value) [3].

$$SNR = \frac{\{E[y_{peak}]\}^2}{Var\{y_{peak}\}} \tag{9}$$

Accordingly, $E[y_{peak}] = E[U_1 + U_2] = 0$ (10)

Equation (10) implies that under the above two assumptions, the SNR becomes zero, which obviously is highly undesired.

B. Fixed Instantaneous Frequency

The least time-varying interference is the one whose frequency is fixed, independent of time. In this case, we take $\omega_n = \omega_0$. The receiver output due to the signal is

$$\begin{aligned}
U_1 &= \sum_{n=1}^L \sum_{m=-1}^1 h(m) p(n-m) = \sum_{n=1}^L (p(n-1) + p(n+1) - 2 \cos(\omega_0) p(n)) p(n) \\
&= \sum_{n=1}^L p(n) p(n-1) + \sum_{n=1}^L p(n) p(n+1) - 2 \sum_{n=1}^L \cos(\omega_0) p^2(n)
\end{aligned} \tag{11}$$

and its expected value is given by

$$E[U_1] = \sum_{n=1}^L E[p(n)p(n-1)] + \sum_{n=1}^L E[p(n)p(n+1)] - 2 \sum_{n=1}^L E[\cos(\omega_0)] = -2L \cos(\omega_0) \quad (12)$$

To obtain the mean square value, we first find

$$\begin{aligned} U_1^2 = & \sum_{n=1}^L \sum_{k=1}^L p(n)p(n-1)p(k)p(k-1) + \sum_{n=1}^L \sum_{k=1}^L p(n)p(n-1)p(k)p(k+1) - 2 \sum_{n=1}^L \sum_{k=1}^L p(n)p(n-1)\cos(\omega_0) \\ & + \sum_{n=1}^L \sum_{k=1}^L p(n)p(n+1)p(k)p(k-1) + \sum_{n=1}^L \sum_{k=1}^L p(n)p(n+1)p(k)p(k+1) - 2 \sum_{n=1}^L \sum_{k=1}^L p(n)p(n+1)\cos(\omega_0) \\ & - 2 \sum_{n=1}^L \sum_{k=1}^L p(k)p(k-1)\cos(\omega_0) - 2 \sum_{n=1}^L \sum_{k=1}^L p(k)p(k-1)\cos(\omega_0) + 4 \sum_{n=1}^L \sum_{k=1}^L \cos^2(\omega_0) \end{aligned} \quad (13)$$

Then, by taking the expected value of (13), the mean-square value and the variance, due to the signal, are respectively given by

$$E[U_1^2] = 4L - 2 + 4L^2 \cos^2(\omega_0), \quad \text{Var}[U_1^2] = 4L - 2 \quad (14)$$

On the other hand, the correlator output due to the noise is

$$U_2 = \sum_{n=1}^L [w(n-1) - 2w(n)\cos\omega_0 + w(n+1)]p(n) \quad (15)$$

whose mean value is zero. It is straightforward to show that the square value of the output noise is given by the following expression

$$\begin{aligned} U_2^2 = & \sum_{n=1}^L \sum_{k=1}^L p(n)w(n-1)w(k-1)p(k) + \sum_{n=1}^L \sum_{k=1}^L p(n)w(n-1)w(k+1)p(k) \\ & - 2 \sum_{n=1}^L \sum_{k=1}^L p(n)w(n-1)w(k)p(k)\cos(\omega_0) + \sum_{n=1}^L \sum_{k=1}^L p(n)w(n+1)p(k)w(k-1) \\ & + \sum_{n=1}^L \sum_{k=1}^L p(n)w(n+1)p(k)w(k+1) - 2 \sum_{n=1}^L \sum_{k=1}^L p(n)w(n+1)w(k)p(k)\cos(\omega_0) \\ & - 2 \sum_{n=1}^L \sum_{k=1}^L w(n)p(n)w(k-1)p(k)\cos(\omega_0) - 2 \sum_{n=1}^L \sum_{k=1}^L w(n)p(n)w(k+1)p(k)\cos(\omega_0) \\ & + 4 \sum_{n=1}^L \sum_{k=1}^L w(n)p(n)w(k)p(k)\cos^2(\omega_0) \end{aligned} \quad (16)$$

Due to the zero mean property, the mean square value of the noise term at the correlator output is equal to its variance and is given by

$$E[U_2^2] = L\sigma^2 + L\sigma^2 + 4L\sigma^2 \cos^2(\omega_0) = 2L\sigma^2(1 + 2\cos^2(\omega_0)) \quad (17)$$

From (16) and (17), the correlator output signal-to-noise ratio is

$$SNR = \frac{4L^2 \cos^2(\omega_o)}{2L + 2L\sigma^2(1 + 2\cos^2(\omega_o))} = \frac{2L \cos^2(\omega_o)}{1 + \sigma^2(1 + 2\cos^2(\omega_o))} \quad (18)$$

According to (18), the correlator output SNR is symmetric around $\omega_o = \pi/2$, as shown in Fig.8. It is evident from this Figure that the performance is highly dependent on frequency. It reaches a minimum (SNR=0) at $\omega_o = \pi/2$, where it exhibits the poorest performance. The highest performance is achieved at zero frequency, irrespective of the noise power.

Case II Five Coefficient Excision Filters

Because of equations (10) and (18), the performance of the three-coefficient filter under both extreme cases of time-varying interference is unsatisfactory. For this reason, we propose below a five coefficient filter which places a second order, instead of a first order, notch at the interference instantaneous frequency. We show that this filter will mitigate both problems which arise in Case I. The five coefficient linear phase FIR filter is

$$h(n) = [1 \quad -4\cos(\omega_n) \quad 2 + 4\cos^2(\omega_n) \quad -4\cos(\omega_n) \quad 1] \quad (19)$$

which is obtained by convolving the impulse response of the three-coefficient filter in Case I with itself.

A. Randomly Changing IF

It is clear that the middle term of the filter impulse response in (19) is no longer a cosinusoidal function, but rather includes the square value of the cosinusoid. As a result, the expected value of the correlator output under randomly changing IF will no longer be zero, forbidding the SNR to be zero, as in Case 1. Using the new filter,

$$U_1 = \sum_{n=1}^L [p(n-2) - 4p(n-1)\cos(\omega_n) + (2 + 4\cos^2(\omega_n))p(n) - 4p(n+1)\cos(\omega_n) + p(n+2)]p(n) \quad (20)$$

The expected values of the first, second, forth, and fifth terms in the above summation are all zeros due to the uncorrelatedness of the PN and the interference instantaneous frequency as well as the zero mean property of both the PN sequence and the cosinusoidal function. The expected value of the middle term is given by

$$E[U_1] = \sum_{n=1}^L E[2 + 4 \cos^2(\omega_n)] = 2L + \frac{4}{2\pi} \int_0^{2\pi} \cos^2(\omega_n) d\omega_n = 4L \quad (21)$$

Because the noise is uncorrelated with the PN sequence, $E[U_2] = 0$. Therefore the numerator in equation (9) is given by

$$E[y_{peak}] = E^2[U_1] = 16L^2 \quad (22)$$

The variance of U can be computed as

$$\begin{aligned} E[U_1^2] = & \sum_{k=1}^L \sum_{n=1}^L E\{p(n)p(n-2)p(k)p(k-2)\} + \sum_{k=1}^L \sum_{n=1}^L E\{p(n)p(n+2)p(k)p(k+2)\} \\ & + 16 \sum_{k=1}^L \sum_{n=1}^L E\{p(n)p(n-1)p(k)p(k-1)\cos(\omega_n)\cos(\omega_k)\} \\ & + 16 \sum_{k=1}^L \sum_{n=1}^L E\{p(n)p(n+1)p(k)p(k+1)\cos(\omega_n)\cos(\omega_k)\} \\ & + \sum_{k=1}^L \sum_{n=1}^L E\{(2 + 4\cos(\omega_n))(2 + 4\cos(\omega_k))\} \\ & + \sum_{k=1}^L \sum_{n=1}^L p(n)p(n-2)p(k)p(k+2) + \sum_{k=1}^L \sum_{n=1}^L p(k)p(k-2)p(n)p(n+2) \end{aligned} \quad (23)$$

The rest of the terms in the mean square value of U_1 represent the cross products of the individual terms. These products are zeros, due to the PN sequence and the instantaneous frequency characteristics. The only non-zero terms in the first and second double summations of equation (23) are those corresponding to $n=k$. These terms take unit values. As such, the first two double summations are simply L . Similarly, the third and forth double summations take the value $8L$. In the last two double summations, there are $L-2$ nonzero terms corresponding to $n=k+2$ and $n=k-2$, respectively. The fifth double summation can be simplified to

$$\begin{aligned}
& \sum_{n=1}^L \sum_{k=1}^L E\{ (4 + 8 \cos^2(\omega_n) + 8 \cos^2(\omega_k) + 16 \cos^2(\omega_n) \cos^2(\omega_k)) \} \\
& = 4L^2 + 4L^2 + 4L^2 + 16 \sum_{n=1}^L E\{ (0.5(1 + 2 \cos(2\omega_n))^2) \} + 16 \sum_{n=1}^L \sum_{k \neq n, k=1}^L \cos^2(\omega_n) \cos^2(\omega_k) \\
& = 12L^2 + 4 \sum_{n=1}^L E\{ (1 + 2 \cos(2\omega_n) + \cos^2(\omega_n)) \} + 4(L^2 - L) = 12L^2 + 6L + 4L^2 - 2L = 16L^2 + 2L
\end{aligned} \tag{24}$$

By adding the values of the different terms constituting $E[U_1^2]$, we obtain

$$E[U_1^2] = 16L^2 + 22L - 2 \tag{25}$$

Now we focus on the mean square value of the correlator output due to noise, which is denoted by U_2 ,

$$\begin{aligned}
U_2 = \sum_{n=1}^L \{ & w(n-2) - 4w(n-1)\cos(\omega_n) + (2 + 4\cos^2(\omega_n))w(n) \\
& - 4w(n+1)\cos(\omega_n) + w(n+2) \} p(n)
\end{aligned} \tag{26}$$

Applying the expectation operator to the square value of the above equation, we obtain

$$\begin{aligned}
E[U_2^2] = & \sum_{k=1}^L \sum_{n=1}^L E\{ p(n)w(n-2)p(k)w(k-2) \} + \sum_{k=1}^L \sum_{n=1}^L E\{ p(n)w(n+2)p(k)w(k+2) \} \\
& + 16 \sum_{k=1}^L \sum_{n=1}^L E\{ p(n)w(n-1)p(k)w(k-1)\cos(\omega_n)\cos(\omega_k) \} \\
& + 16 \sum_{k=1}^L \sum_{n=1}^L E\{ p(n)w(n+1)p(k)w(k+1)\cos(\omega_n)\cos(\omega_k) \} \\
& + \sum_{k=1}^L \sum_{n=1}^L E\{ (2 + 4\cos(\omega_n))(2 + 4\cos(\omega_k))w(n)p(n)w(k)p(k) \}
\end{aligned} \tag{27}$$

Each of the first two double summations yields $L\sigma^2$. The value of the third and the forth double summations is $8L\sigma^2$. The last double summation is $14L\sigma^2$. All of the above values were obtained based on the assumption that the PN sequence and the white noise sequence are uncorrelated, and the noise is a white random stationary process. Since

$$E[U_1] = 16L^2, \quad E[U_2] = 0, \tag{28}$$

then

$$\text{var}[U_1] = 22L - 2, \quad \text{var}[U_2] = 32L\sigma^2 \tag{29}$$

Substituting (28) and (29) in (9), we obtain

$$SNR_o = \frac{16L^2}{22L - 2 + 32L\sigma^2} \quad (30)$$

For reasonably high value of L, the above expression can be simplified to

$$SNR_o = \frac{L}{\frac{22}{16} + 2\sigma^2} \quad (31)$$

If there is no suppression filter (preprocessor disabled), then the corresponding output SNR is

$$SNR_{no} = \frac{L}{\rho^2 + \sigma^2} \quad (32)$$

From equations (31) and (32), it is evident that in a jammer free environment, the five-coefficient excision filter reduces the gain by approximately 3 dB from the case where no filter is applied. This is due to the excision filter frequency characteristics which remove part of the signal in the vicinity of the notch frequency. However, in the presence of an interference, the excision filter improves the SNR by approximately $10\log(\rho^2/2)$, where ρ^2 is the interference power.

B. Fixed Instantaneous Frequency

In this case, the output of the correlator due to the signal is

$$U_1 = \sum_{n=1}^L [p(n-2) - 4p(n-1)\cos(\omega_o) + (2 + 4\cos^2(\omega_o))p(n) - 4p(n+1)\cos(\omega_o) + p(n+2)]p(n) \quad (33)$$

and its mean value

$$E[U_1] = \sum_{n=1}^L E[2 + 4\cos^2(\omega_n)] = 2L + 4L\cos^2 \omega_o \quad (34)$$

Because the noise is uncorrelated with the PN sequence, $E[U_1] = 0$. The numerator in equation (9) is, therefore, given by

$$E[y_{peak}] = E^2[U_1] = 4(L + 2L\cos^2 \omega_o)^2 \quad (35)$$

The variance of U_1 can be computed as

$$\begin{aligned}
E[U_1^2] = & \sum_{k=1}^L \sum_{n=1}^L E\{p(n)p(n-2)p(k)p(k-2)\} + \sum_{k=1}^L \sum_{n=1}^L E\{p(n)p(n+2)p(k)p(k+2)\} \\
& + 16 \sum_{k=1}^L \sum_{n=1}^L E\{p(n)p(n-1)p(k)p(k-1)\cos^2(\omega_o)\} + 16 \sum_{k=1}^L \sum_{n=1}^L E\{p(n)p(n+1)p(k)p(k+1)\cos^2(\omega_o)\} \\
& + \sum_{k=1}^L \sum_{n=1}^L E\{(2+4\cos(\omega_o))^2\} + \sum_{k=1}^L \sum_{n=1}^L p(n)p(n-2)p(k)p(k+2) + \sum_{k=1}^L \sum_{n=1}^L p(k)p(k-2)p(n)p(n+2) \\
& + 32 \sum_{k=1}^L \sum_{n=1}^L p(n-1)p(n)p(k+1)p(k)\cos^2(\omega_o)
\end{aligned} \tag{36}$$

which simplifies to

$$E[U_1^2] = 4L - 4 + 64L\cos^2(\omega_o) + L^2(2 + 4\cos^2(\omega_o))^2 - 32\cos^2(\omega_o) \tag{37}$$

On the other hand, the correlator output due to noise, U_2 , is given by

$$U_2 = \sum_{n=1}^L [w(n-2) - 4w(n-1)\cos(\omega_n) + (2 + 4\cos^2(\omega_n))w(n) - 4w(n+1)\cos(\omega_n) + w(n+2)]p(n)$$

Applying the expectation operator to the square value of the above equation, we obtain

$$\begin{aligned}
E[U_2^2] = & \sum_{k=1}^L \sum_{n=1}^L E\{p(n)w(n-2)p(k)w(k-2)\} + \sum_{k=1}^L \sum_{n=1}^L E\{p(n)w(n+2)p(k)w(k+2)\} \\
& + 16 \sum_{k=1}^L \sum_{n=1}^L E\{p(n)w(n-1)p(k)w(k-1)\cos^2(\omega_o)\} + 16 \sum_{k=1}^L \sum_{n=1}^L E\{p(n)w(n+1)p(k)w(k+1)\cos^2(\omega_o)\} \\
& + \sum_{k=1}^L \sum_{n=1}^L E\{(2 + 4\cos(\omega_n))^2 w(n)p(n)w(k)p(k)\}
\end{aligned} \tag{38}$$

Accordingly,

$$\begin{aligned}
\text{var}(U_1) &= 4L - 4 + 64L\cos^2(\omega_o) - 32\cos^2(\omega_o) \\
\text{var}(U_2) &= 2L\sigma^2 + 32L\sigma^2\cos^2(\omega_o) + L\sigma^2(2 + 4\cos^2(\omega_o))
\end{aligned} \tag{39}$$

Substituting (33) and (39) in (9), we obtain

$$SNR_o = \frac{4L(1 + 2\cos^2(\omega_o))^2}{4 - 4/L + 64\cos^2(\omega_o) - (32/L)\cos^2(\omega_o) + 2\sigma^2 + 32\sigma^2\cos^2(\omega_o) + \sigma^2(2 + 4\cos^2(\omega_o))^2}, \tag{40}$$

For reasonably high value of L, the above expression can be simplified to

$$SNR_o = \frac{2L(1 + 2\cos^2(\omega_o))^2}{2 + 32\cos^2(\omega_o) + \sigma^2 + 16\sigma^2\cos^2(\omega_o) + 2\sigma^2(1 + 2\cos^2(\omega_o))^2}, \tag{41}$$

In contrast with Case I of the three coefficient excision filter, the use of five coefficients filter yields the highest performance at $\omega_o = \pi/2$, as shown in Fig.8. Further, Case II has a clear advantage over Case I in that the excision filter performance seems constant across over the Nyquist interval, with an exception at and in the vicinity of $\omega_o = \pi/2$.

Case III: Two Three-Coefficient Excision Filters

A. Randomly Changing IF

Next, we apply 3-coefficient filter to both the input sequence and the receiver PN sequence. In this case, the correlator output is

$$U = \sum_{n=1}^L p_o(n)p_o(n) + \sum_{n=1}^L i_o(n)p_o(n) + \sum_{n=1}^L p_o(n)w(n) \quad (42)$$

Assuming that the interference is completely excised, then the second term in the above summation is zero. The mean value of the correlator output is

$$E[U] = E[U_1] + E[U_2] \quad (43)$$

where the first term is due to the signal and the second term is due to the noise. It is easily shown that

$$E[U_1] = E\left[\sum_{n=1}^L (p(n-1) + p(n+1) - 2\cos(\omega_n)p(n))(p(n-1) + p(n+1) - 2\cos(\omega_n)p(n))\right] \quad (44)$$

$$E[U_2] = E\left[\sum_{n=1}^L (w(n-1) + w(n+1) - 2\cos(\omega_n)w(n))(p(n-1) + p(n+1) - 2\cos(\omega_n)p(n))\right] \quad (45)$$

Using the uncorrelatedness of the PN sequence and the noise sequences, the mean value of the correlator output signal is $4L$, where as the correlator output noise is zero. Further,

$$\begin{aligned} E[U_1^2] &= \sum_{k=1}^L \sum_{n=1}^L 4 + 4 \sum_{k=1}^L \sum_{n=1}^L E\{p(n-1)p(n+1)p(k-1)p(k+1)\} \\ &\quad + 16 \sum_{k=1}^L \sum_{n=1}^L E\{p(n)p(n+1)p(k)p(k+1)\cos(\omega_n)\cos(\omega_k)\} \\ &\quad + 16 \sum_{k=1}^L \sum_{n=1}^L E\{p(n)p(n-1)p(k)p(k-1)\cos(\omega_n)\cos(\omega_k)\} \\ &\quad + 16 \sum_{n=1}^L \sum_{k=1}^L E\{\cos^2(\omega_n)\cos^2(\omega_k)\} + 16 \sum_{n=1}^L \sum_{k=1}^L E\{\cos^2(\omega_n)\} \\ &= 4L^2 + 4L + 8L + 8L + 4L^2 + 2L + 8L^2 = 16L^2 + 22L \end{aligned} \quad (46)$$

Similarly, it can readily be shown that

$$E[U_2^2] = 20L\sigma^2 \quad (47)$$

$$\text{So, } SNR_o = \frac{16L^2}{22L - 2 + 22L\sigma^2} \quad (48)$$

For a reasonably high value of L, the above expression can be approximated by

$$SNR_o = \frac{L}{\frac{22}{16} + \frac{20}{16}\sigma^2} \quad (49)$$

Comparing Case II and Case III, it is clear that the signal to noise ratio improves if we process the PN sequence at the receiver with the same Excision filter.

B. Fixed Instantaneous Frequency

In this case, the mean values of the signal and noise components of the correlator output are

$$\begin{aligned} E[U_1] &= E\left[\sum_{n=1}^L (p(n-1) + p(n+1) - 2\cos(\omega_n)p(n))(p(n-1) + p(n+1) - 2\cos(\omega_n)p(n))\right] \\ E[U_2] &= E\left[\sum_{n=1}^L (w(n-1) + w(n+1) - 2\cos(\omega_n)w(n))(p(n-1) + p(n+1) - 2\cos(\omega_n)p(n))\right] \\ E[U_1] &= E\left[\sum_{n=1}^L p^2(n-1) + p^2(n+1) + 2p(n-1)p(n+1) - 4\cos(\omega_n)p(n)p(n-1) \right. \\ &\quad \left. - 4\cos(\omega_n)p(n)p(n+1) + 4\cos^2(\omega_n)p^2(n)\right] = 2L + 4L\cos^2(\omega_o) \end{aligned} \quad (50)$$

It is easily shown that by using the uncorrelatedness of the PN sequence and the noise sequences, the mean value of the correlator output signal is 4L, whereas the correlator output noise is zero. The mean value of U_1 is given by

$$\begin{aligned} E[U_1^2] &= \sum_{k=1}^L \sum_{n=1}^L 4 + 4 \sum_{k=1}^L \sum_{n=1}^L E\{p(n-1)p(n+1)p(k-1)p(k+1)\} \\ &+ 16 \sum_{k=1}^L \sum_{n=1}^L E\{p(n)p(n+1)p(k)p(k+1)\cos^2(\omega_o)\} + 16 \sum_{k=1}^L \sum_{n=1}^L E\{p(n)p(n-1)p(k)p(k-1)\cos^2(\omega_o)\} \\ &+ 16 \sum_{k=1}^L \sum_{n=1}^L E\{p(n)p(n+1)p(k)p(k-1)\cos^2(\omega_o)\} + 16 \sum_{k=1}^L \sum_{n=1}^L E\{p(n)p(n-1)p(k)p(k+1)\cos^2(\omega_o)\} \\ &+ 16 \sum_{n=1}^L \sum_{k=1}^L E\{\cos^4(\omega_o)\} + 16 \sum_{n=1}^L \sum_{k=1}^L E\{\cos^2(\omega_o)\} = 4L^2(1 + 4\cos^2(\omega_o) + 4\cos^4(\omega_o)) + 4L(1 + 8\cos^2(\omega_o)) \\ &\quad + 32(L-1)\cos^2(\omega_o) \end{aligned}$$

$$\text{var}[U_1] = 64L\cos^2(\omega_o) - 32\cos^2(\omega_o) + 4L \quad (51)$$

Similarly, it can be shown that

$$\begin{aligned} E[U_2^2] &= 4L\sigma^2 + 16L\sigma^2 \cos^2(\omega_o) + 16L\sigma^2 \cos^4(\omega_o) + 2(L-2)\sigma^2 + 64(L-1)\sigma^2 \cos^2(\omega_o) \\ &= 4L\sigma^2(1 + 2\cos^2(\omega_o))^2 + 2(L-2)\sigma^2 + 64(L-1)\sigma^2 \cos^2(\omega_o) \end{aligned} \quad (52)$$

$$SNR_o = \frac{4(1 + 2\cos^2(\omega_o))^2 L^2}{64L\cos^2(\omega_o) - 32\cos^2(\omega_o) + 4L + 4L\sigma^2(1 + 2\cos^2(\omega_o))^2 + 2(L-2)\sigma^2 + 64(L-1)\sigma^2 \cos^2(\omega_o)}$$

For large values of L, the above expression can be approximated to

$$SNR_o = \frac{2(1 + 2\cos^2(\omega_o))^2 L}{32(1 + \sigma^2)\cos^2(\omega_o) + 2 + 2\sigma^2(1 + 2\cos^2(\omega_o))^2 + \sigma^2} \quad (53)$$

Similar to Case II, Case III shows peak performance at $\omega_o = .5\pi$ and almost equal performance over $[0, \pi]$. Fig.8 depicts the SNR for the above three cases with the noise power $\sigma^2 = .01, 1, 1000$, respectively. It is evident that Case III outperforms Case II, specifically at high noise power, where averaging over filter taps tend to amplify the noise contribution at the output. Case I favors very low and very high frequencies and in this sense has a stopband filter characteristics. Cases II and III favor mid frequencies and as such each possesses a bandpass filter characteristics.

Simulations

In the two examples presented below, we show the effect of the open loop adaptive filtering used in the interference excision on a binary phase shift keying (BPSK) waveform. Both chirp and frequency hopping interfering signals with Jammer-to-signal (J/S) ratio of 20 dB are considered. A BPSK waveform is generated over 100 samples and taken as the desired signal with pulse width(chip length) of 3 data samples.

In the first example, the interference is taken as a chirp signal, shown in Fig.9(a). The instantaneous frequency of the chirp is shown in Fig. 9(b). Using the Wigner distribution kernel, the instantaneous frequency is estimated from the time-frequency distribution and then used to design a three coefficient filter open loop adaptive filter. Fig. 9(c,d,e) shows the

interference and its spectrum before and after processing by the filter. It is evident that the interference is almost annihilated with a reduction of at least -30 dB of its power. Next we examine the effect of the adaptive filter on the desired signal. Fig. 9(f) shows the normalized BPSK sequence before and after the application of the filter. The noticeable increase in the amplitude is due to the fact that a constant K processed by a LTI FIR filter whose impulse response is $[1 \ -2\cos \omega_n \ 1]$ increases in value for increased ω_n (instantaneous frequency of the chirp). On the other hand, the spikes exhibited in the desired signal output at the beginning and the end of each pulse are due to the changing of the sample values within the filter at the time instants of sudden transitions. These spikes can be easily removed by clipping or only considering the center point in each pulse (see the second example).

In the second example, the interference is taken as a frequency hopping signal, shown in Fig.10(a). The instantaneous frequency of the interference is shown in Fig. 10(b). Figs 10(c,d,e,f,g) are the counterparts of those in the above example. These figures are obtained by using the Born Jordan kernel [8]. We maintain that the abrupt changes in the filter output due to the interference at the beginning of each hop is due to the fact that the data samples within the filter at these time instants do not belong to the same sinusoid. These undesired spikes can be easily detected and either clipped or entirely removed by declaring these data samples as highly contaminated by the interference.

Test results for four specific cases can be seen in Figures 11 and 12 are dedicated to the Bit Error Rate (BER) curves. For the two types of jammers, a baseline was established for the direct sequence spread spectrum receiver (preprocessor disabled). The jammer power was varied from 10 to 60 dB (relative to the signal energy) in increments of 5 dB. The swept jammer case (chirp signal) is shown in Fig. 11. In this case, the BER was .0544 for a 10 dB J/S ratio. For the case of 20 dB J/S, the BER increased to .430. With the preprocessor enabled, the BER for 10 dB J/S was .0309. For the highest jammer power (60 dB) the BER

was .291, demonstrating that with the excision filter enabled, the receiver produced fewer error than the case with the filter disabled and only a 20 dB J/S.

The results for the hopped jammer case were similar, as can be seen in Fig. 12. With the preprocessor disabled and a J/S of 10 dB, the BER was .0569. This increased to 0.429 when the J/S ratio was increased to 20 dB. With the preprocessor enabled, the BER for 10 dB J/S was .00776. For the case of 60 dB J/S, the BER was .337 which, again, is lower than the BER without the preprocessor enabled and a J/S ratio of 20 dB.

Further Development

Out of this contract, the following papers have either been published or accepted for publications (The papers marked by * indicate joint collaboration with Rome Lab technical staff members). We encourage the reader of this report to consult these references for more discussions on the subject.

- Journals

M. Amin, "Recursive Kernels for Time-Frequency Signal Representations," IEEE Signal Processing Letters., January 1996

M. Amin, G. Venkatesan, and J. Carroll, "A Constrained Weighted Least Squares Approach for Time-Frequency Kernel Design," IEEE Trans. on Signal Processing, May 1996.

M. Amin and K. Feng, "Short-Time Fourier Transform Using Cascade Filter Structures," IEEE Transactions on Circuits and Systems, October 1995.

G. Zalubas and M. Amin, "Time-Frequency Kernel Design by the Two-Dimensional Frequency Transformation Method," IEEE Transactions on Signal Processing, September 1995.

* S. Tyler and M. Amin, "Mitigating interference in direct sequence spread spectrum communication systems," Rome Lab. Technical Journal, 1st Volume, June 1995.

- Conferences

- * M. Amin, G. Venkatesan, and J. Patti " Time-frequency distribution kernels using optimum FIR filter design techniques," Proceedings of the Asilomar Conference Signals, Systems, and Computers, Pacific Grove, CA, November 1995.

- * S. Roberts and M. Amin " Linear Vs. Bilinear time-frequency methods for interference mitigation in direct sequence spread spectrum communication systems," Proceedings of the Asilomar Conference Signals, Systems, and Computers, Pacific Grove, CA, November 1995.

- * M. Amin, G. Venkatesan and S. Tyler, " A new approach for spread spectrum using time-frequency distributions," Proceedings of the SPIE Conference on Advanced algorithms and Architectures for Signal Processing, San Diego, CA , July 1995.

- R. Perry and M. Amin," Field programmable gate array implementation of a class of computationally efficient recursive time-frequency distribution kernels," Proceedings of the International Conference on Signal Processing Applications and Technology, Boston, MA, October 1995.

- * J. Patti, S. Roberts, and M. Amin, "Adaptive and block excisions in spread spectrum communication systems using the wavelet transform," Proceedings of the Asilomar Conference on Circuits, Systems, and Computers, Pacific Grove, CA, Oct. 1994.

- * K. Feng and M. Amin, and S. Tyler, "Analysis of recursive multiple widow spectrograms," Proceedings of the IEEE International Symposium on Time-Frequency and Time-Scale Analysis, Philadelphia, Oct. 1994.

Discussions

Below, we present three key issues for extending the research work on interference excision using time-frequency distributions.

A Algorithms for Instantaneous Frequency Estimates

The interference excision techniques based on the use of time-frequency distributions have shown significant improvement in reducing the probability of error over classical techniques implementing spectral estimation methods and closed loop adaptive filters. The key for improved performance is to gain knowledge of the instantaneous frequency in the time-frequency domain and employ it to design a time-varying notch filter. The notch filter places a zero on a unit circle at the interference instantaneous frequency and as such eliminates a vast majority of its power.

We maintain however that if only the instantaneous frequency is required, there is a variety of other techniques which can be used in place of the time-frequency distributions. To name a few; the phase difference method, the zero crossing method, smoothed phase differencing, short-time Fourier transform, and adaptive estimation of phase polynomial [9]. Some of these methods are much simpler to implement than the TFD and may result in a comparable performance under monocomponent high power interference. Fig.13 shows the instantaneous frequency obtained using the zero-crossing technique applied to linear FM, sinusoidal FM and frequency hopping signals, all of 10 dB SNR. Performance is satisfactory and may prove to be equally efficient to the TFD. However, several of these methods are likely to fail under multicomponent jammer scenarios and low SNR. Fig.14 shows the zero-crossing estimates of individual and combined sinusoidal signals. It is clear that the technique fails substantially under multicomponent signals. Nevertheless, it is important to this research effort to investigate the capability of these techniques for interference excision in spread spectrum systems and compare their performance with that of the TFD-based techniques.

B. Combined Instantaneous Frequency/Bandwidth for Improved Interference

Excision

The interference excision method based on the instantaneous frequency estimate using time-frequency distribution is adequate for a narrow class of jammers, namely those of

constant envelope. This class includes a chirp jammer as well as a frequency hopping jammer. For this class, the signal energy is highly localized around the instantaneous frequency, with minimum spreading. For any given time sample, the signal is viewed as a narrowband component whose frequency is equal to the instantaneous frequency. As such, an open loop adaptive filter can be design to excise this specific frequency.

However, the instantaneous frequency alone does not fully characterize nonstationary signals with non constant envelopes. For these signals, the energy can be widely spread around the instantaneous frequency, and the narrowband assumption may no longer be valid. Accordingly, the open loop notch filter should act on reducing the interference energy over its "Instantaneous bandwidth". The later has been shown to be the second moment of the time-frequency distribution[8].

To illustrate the importance of incorporating the instantaneous bandwidth, we present in Fig. 15 the TFD of Southern Right Whale[8]. It is clear that one signal component has larger instantaneous bandwidth than the other. Applying a notch filter that is only based on the instantaneous frequency is likely to allow more energy of the high frequency component to escape to the output. In order to mitigate this problem, the time-varying filter should incorporate the spread of the signal around its instantaneous frequency. In this case, a stop-band filter should replace the notch filter at instants of large bandwidth. We note that the instantaneous bandwidth may be constant or varying with the instantaneous frequency.

We maintain that the instantaneous bandwidth information can't be extracted from applying most of the methods devised for only instantaneous frequency estimates [9]. Therefore, the TFD remains one of the most effective ways to gain full knowledge of the jammer characteristics.

C. Affine and Hyperbolic TFD Receivers

We maintain that the time-frequency shift-invariance property underlying Cohen's class implies a type of time-frequency analysis where the QTFD's analysis characteristics do not

change with time or frequency. Members of this class are spectrograms, the Wigner distribution, the smoothed pseudo Wigner distribution, the Choi-Williams distribution, the cone-kernel representation and the Born-Jordan distribution. In each of these distributions, all time-frequency points are analyzed with the same time resolution and the same frequency resolution. This is similar to the constant bandwidth analysis achieved by the short time Fourier transform, where the analysis bandwidth does not depend on the analysis time or the analysis frequency. This property can be illustrated by rewriting the TFD in terms of Wigner distribution $W_x(t, f)$

$$C_x(t, f) = \iint \Psi(t - t', f - f') W_x(t', f') dt' df'$$

where Ψ is a smoothing function. Fig.(16-a) depicts the time-frequency invariant smoothing operation affiliated with the above equation. A different but important class of quadratic TFD is the affine class[10],

$$T_x(t, f) = \iint \Psi(f(t - t'), f' / f) W_x(t', f') dt' df'$$

The smoothing offered by the above equation is depicted in Fig. (16-b). It is clear from this figure that different time-frequency points are analyzed by different time-frequency resolutions. Fig. 17 compares the Wigner distribution and its affine smoothing (Bertrand distribution). It is evident that the affine smoothing is more suitable for hyperbolic t-f geometry, where no cross terms are present. It is therefore expected that TFD receivers based on the affine class of QTFD should offer a smaller probability of error than Cohen's class of shift-covariant TFD for a large class of signals.

Conclusions

In this report, mitigation of nonstationary interference in spread spectrum communication systems is achieved based on the interference time-varying characteristics, specifically, its instantaneous frequency. The instantaneous frequency estimate obtained from performing the time-frequency distribution is used to construct a finite impulse response filter

which substantially reduces the interference power with a minimum possible distortion of the desired signal. This two step mechanism for interference excision can be viewed as a case of an open loop adaptive filtering. However, contrary to the existing techniques of close loop self-tuning linear predictors or open loop adaptive filtering based on fast Fourier transforms, the filter coefficients in the proposed technique are obtained via time-varying spectral analysis. Closed form expressions of the improvement of SNR at the receiver correlator output using the t-f synthesis technique and the TFD-based adaptive filtering were derived.

Acknowledgment

The author would like to acknowledge Stephen Tyler and John Patti at Rome Lab for their valuable comments and input on this subject. The author is also grateful for Gopal Venkatesan for generating most of the simulation results included in this report.

References

1. M. K. Simon, et al, Spread Spectrum Communications, Computer Science Press, 1985
2. M. G. Amin, "Interference excision in spread spectrum communication systems using time-frequency distributions," Technical Report, AFOSR, Rome lab, Sept. 1994.
3. J. Ketchum and J. Proakis, " Adaptive algorithms for estimating and suppressing narrow band interference in PN spread spectrum systems", IEEE Transactions on Communications, May 1982.
4. S. Theodoridis, N. Kalouptsidis, J. Proakis and G. Koyas," Interference rejection in PN spread spectrum systems with linear phase FIR filters", IEEE Transactions on Communications, vol. 37, no. 9, September 1989.
5. R. Iltis, J. Ritcey ,and L. Milstein, "Interference rejection in FFH systems using least squares estimation techniques", IEEE Transact. on Communications, vol. 38, Dec. 1990.

6. M. Medley, G. Saulnier, and P. Das, "Applications of the wavelet transform in spread spectrum communications systems", SPIE, Wavelet Applic., Orlando, Florida, April 1994.
7. L. Cohen, Time-Frequency Analysis, Prentice-Hall, Englewood Cliffs, New Jersey, 1995.
9. B. Boashash, "Estimating and Interpreting the instantaneous frequency of a signal -- part 2: algorithms and applications," The Proceedings of the IEEE, April 1992.
10. A. Papandreou, F. Hlawatsch, and G. F Boudreaux-Bartels," The hyperbolic class of quadratic time-frequency representations part I: constant-Q Warping, the hyperbolic paradigm, properties, and members," IEEE Transactions on Signal Processing, December 1993.

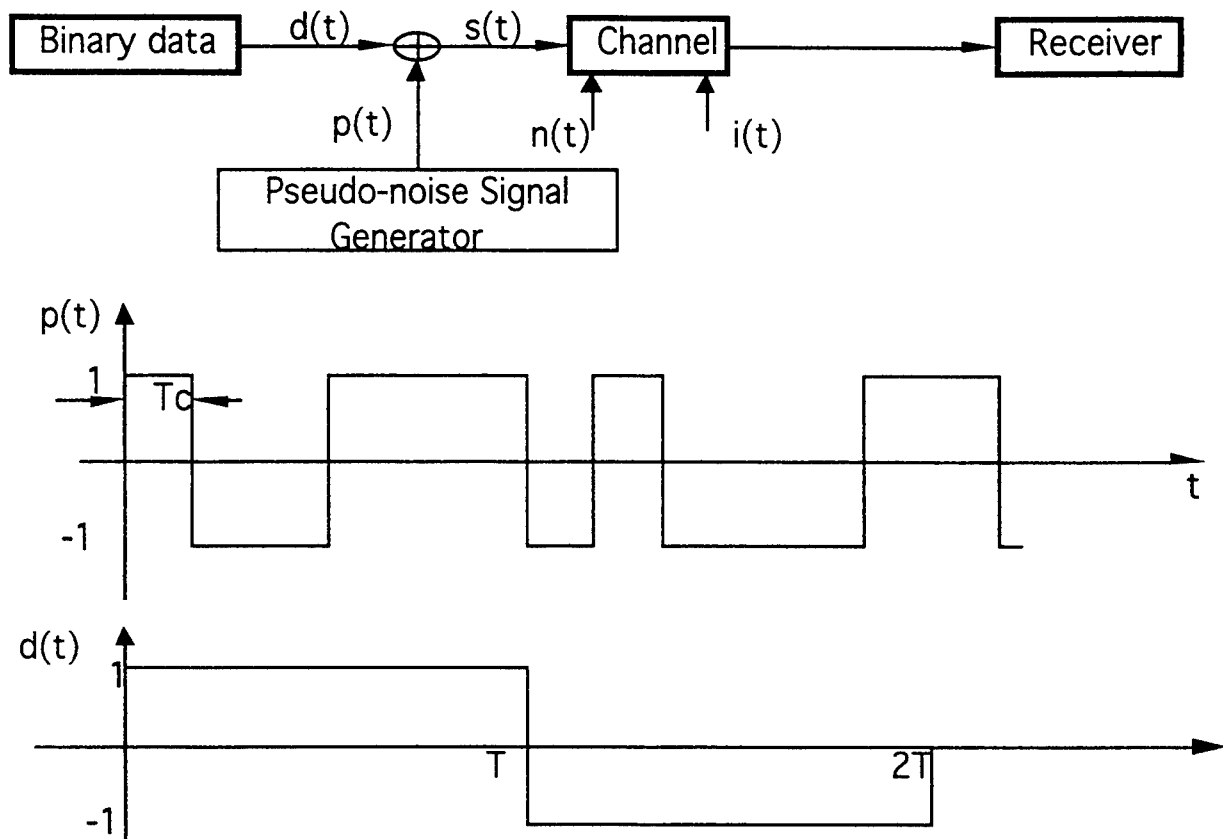


Fig.1 Direct Sequence Spread Spectrum Communication System

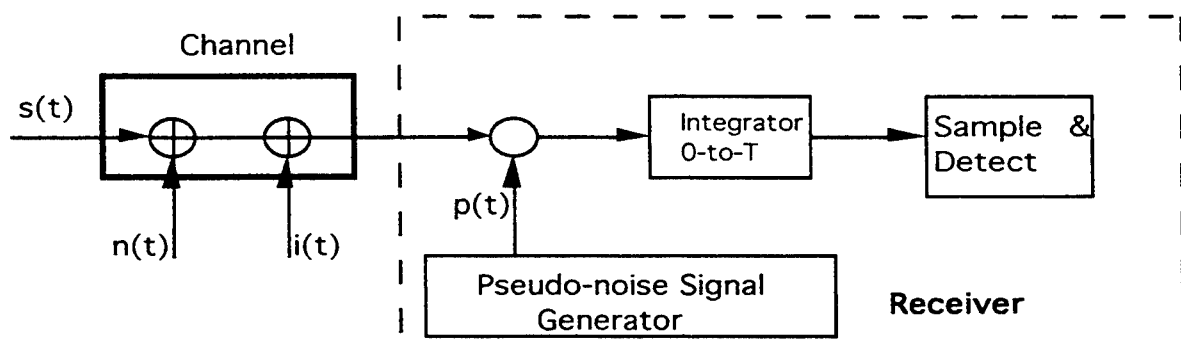


Fig.2 Direct Sequence Spread Spectrum Receiver

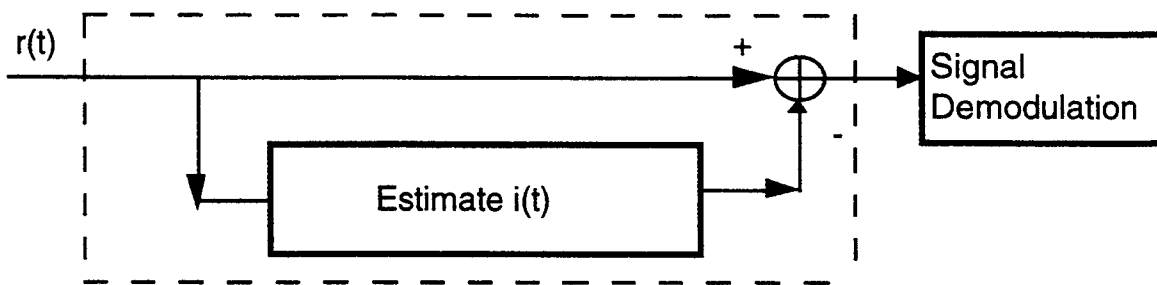


Fig.3 Interference Excision System

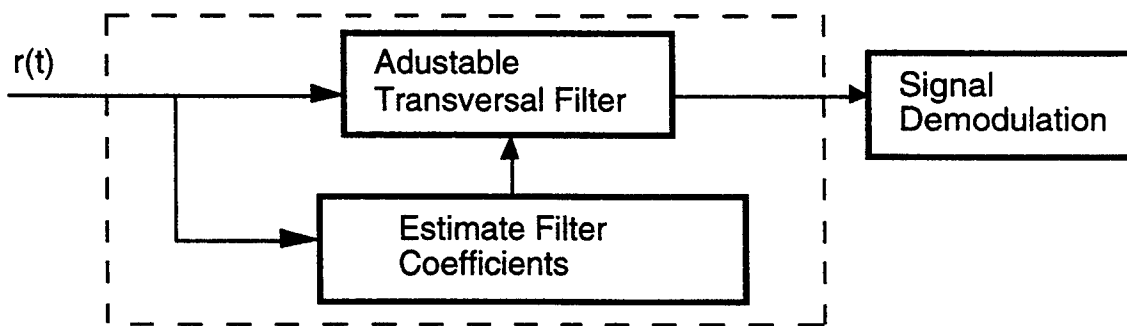


Fig.4 Interference Excision Filter

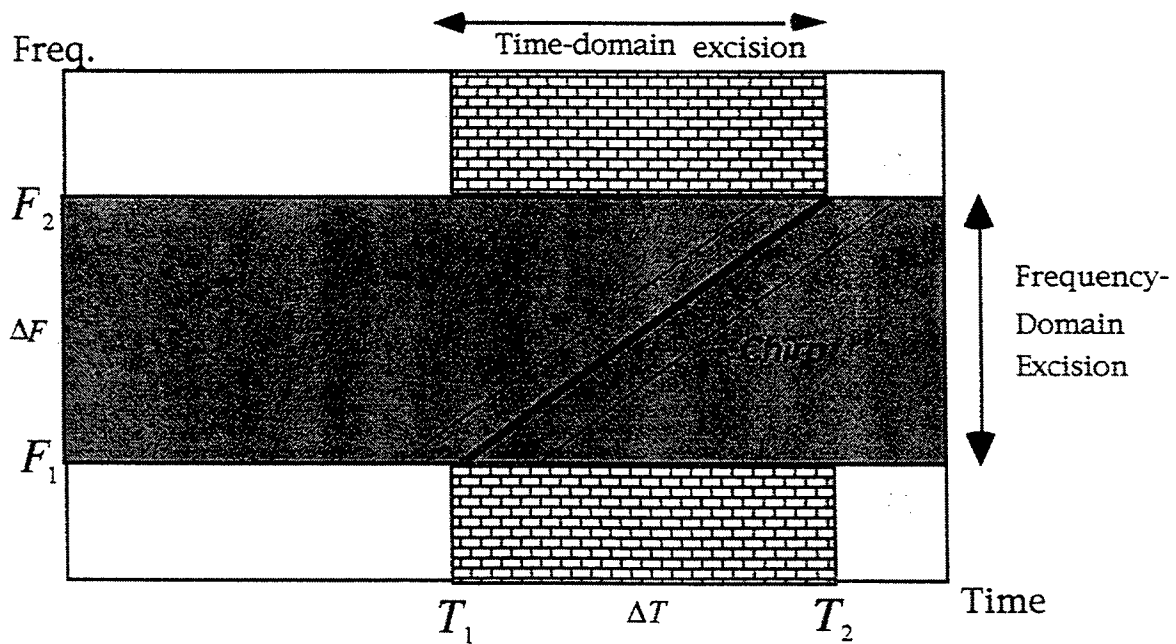
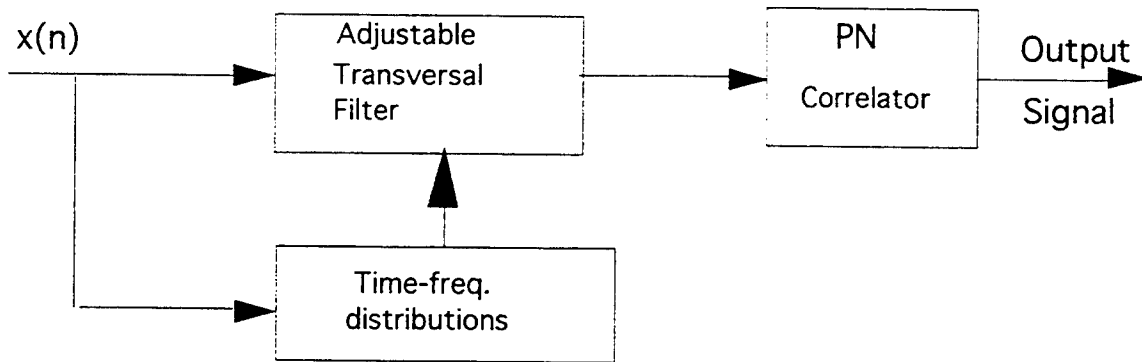
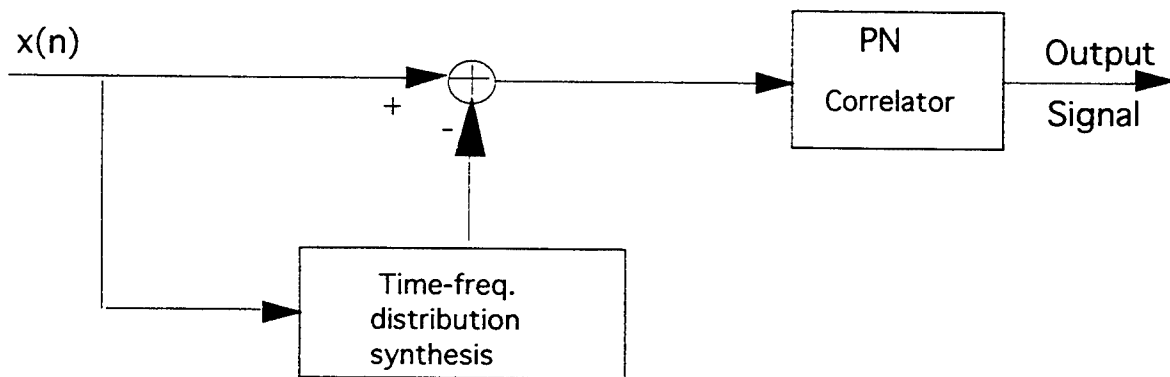


Fig. 5 Excision Methods for Nonstationary signals



(a)



(b)

Fig. 6 Two Proposed Approaches based on TFD (a) Open Loop Adaptive Filtering, (b) Synthesis Method

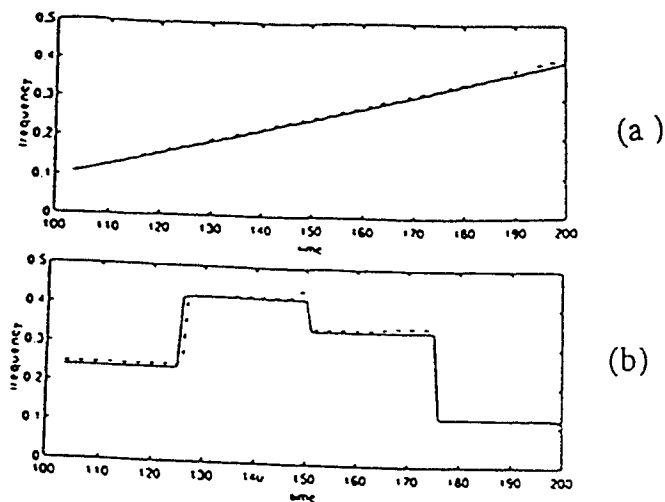


Fig. 7 Instantaneous frequency estimates (a) Chirp/Wigner distribution (b) Frequency hopping/Born Jordan distribution.

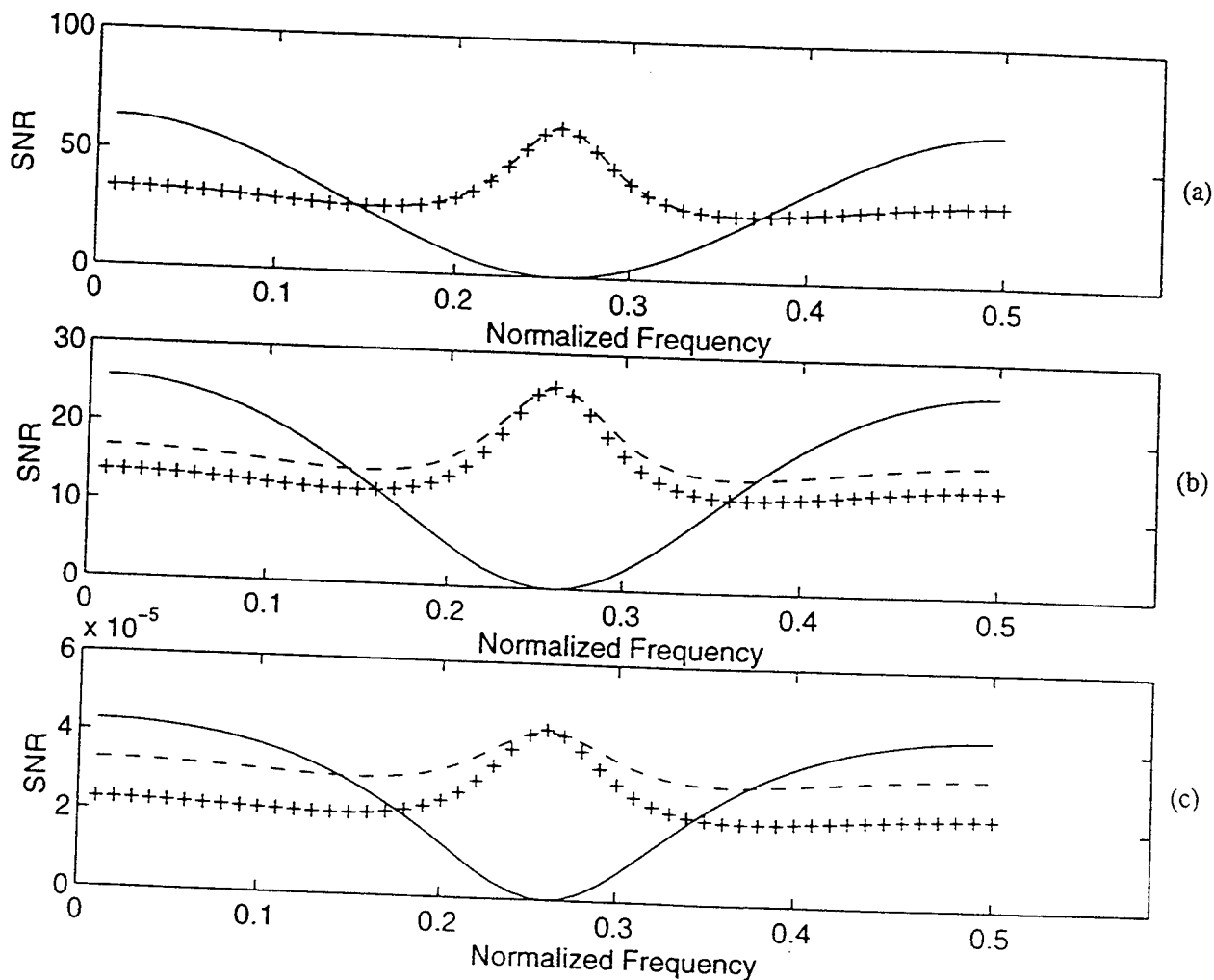


Fig. 8 Excision filters under fixed frequency sinusoids, (-) three coefficient filter (- -) five coefficient filter (+) two three coefficient filter (a) no noise case, (b) SNR is 0 dB, (c) SNR is 30 dB.

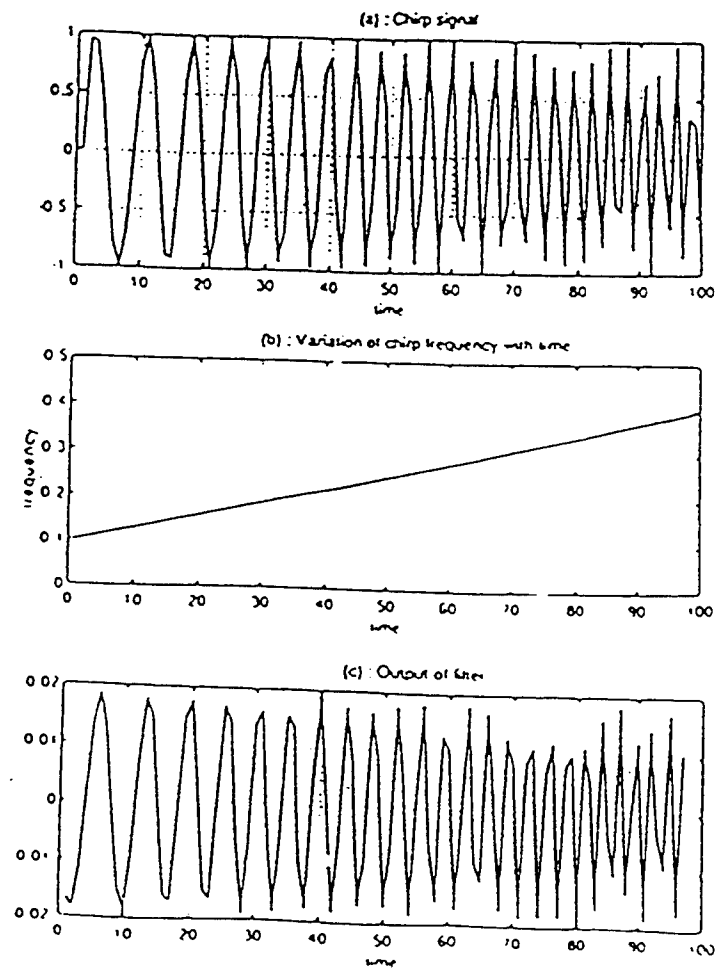


Fig. 9

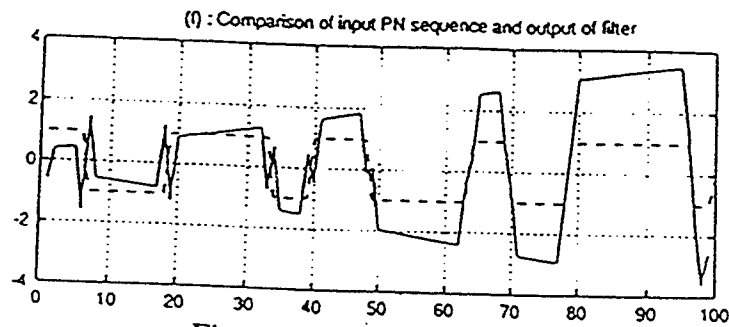
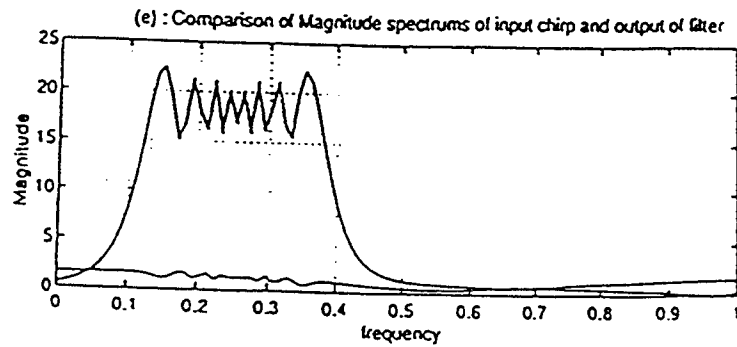
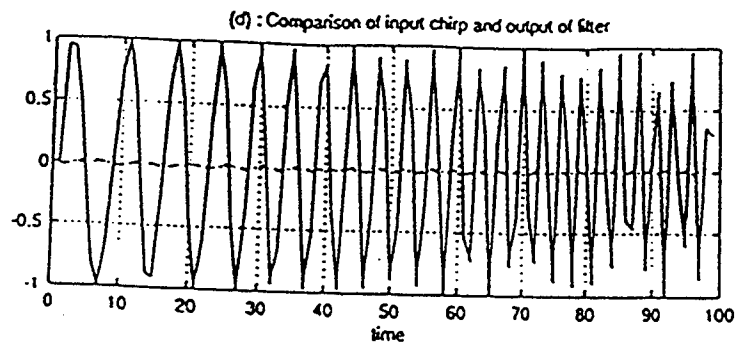
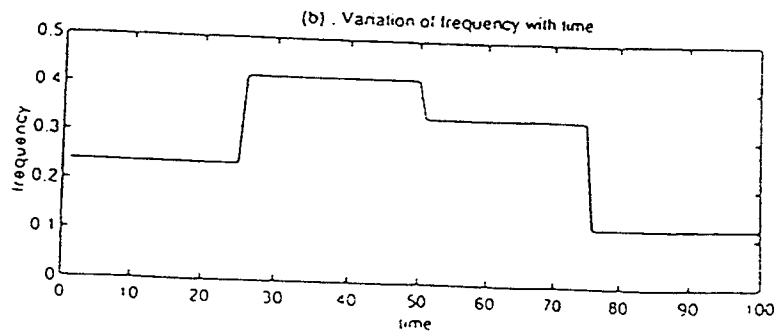
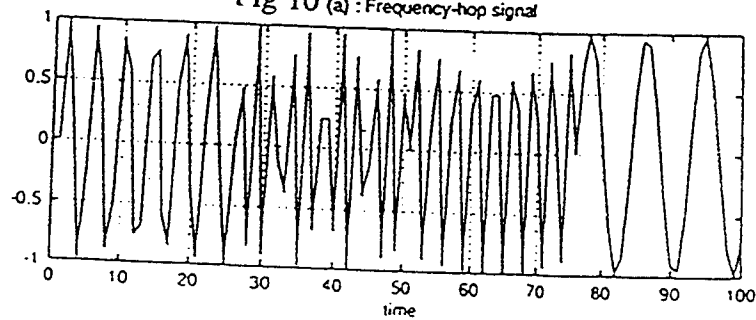


Fig 10 (a) : Frequency-hop signal



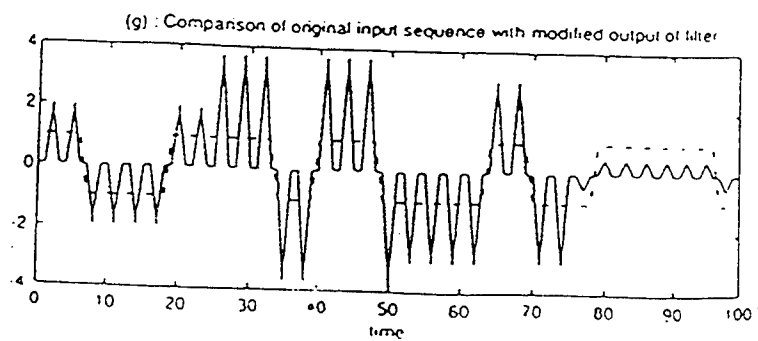
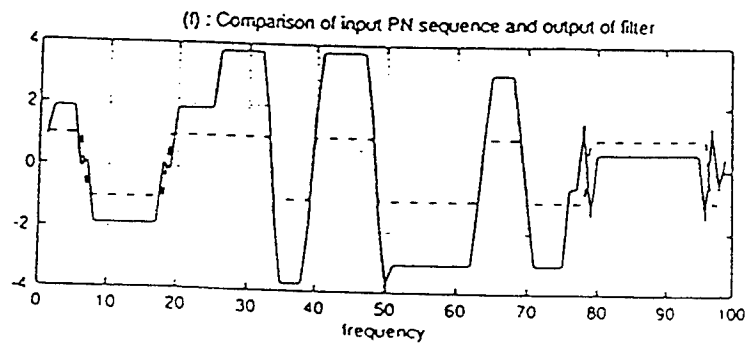
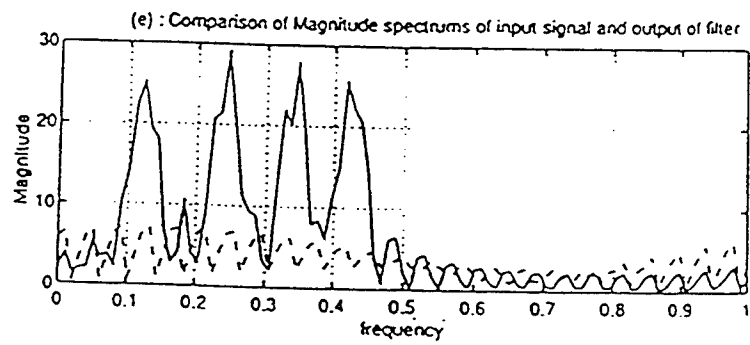
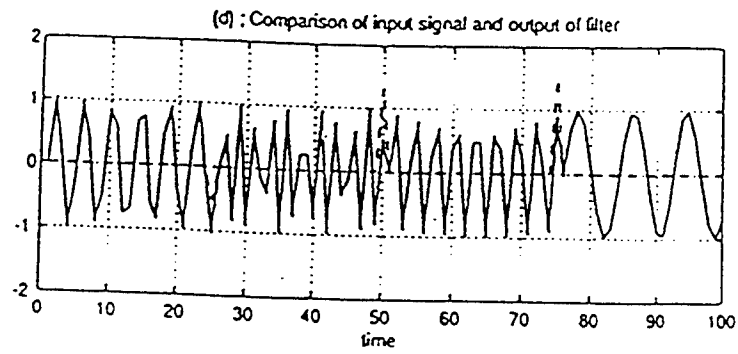
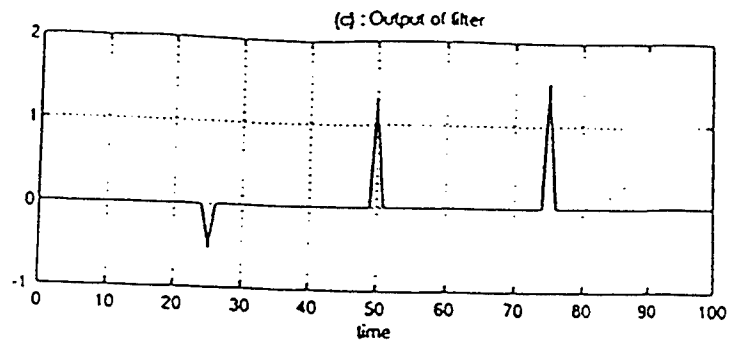


Fig. 11.

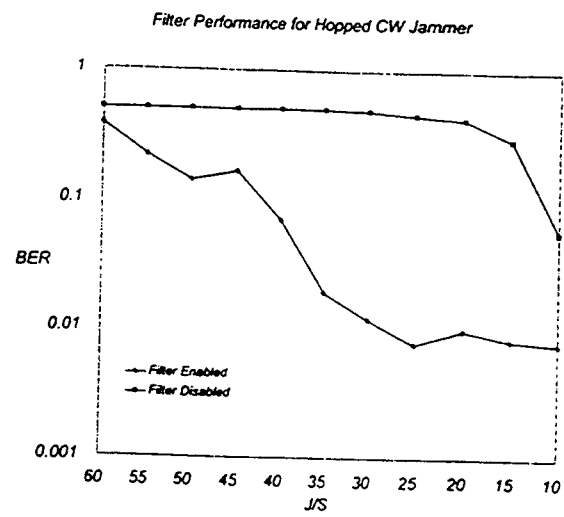


Fig. 12

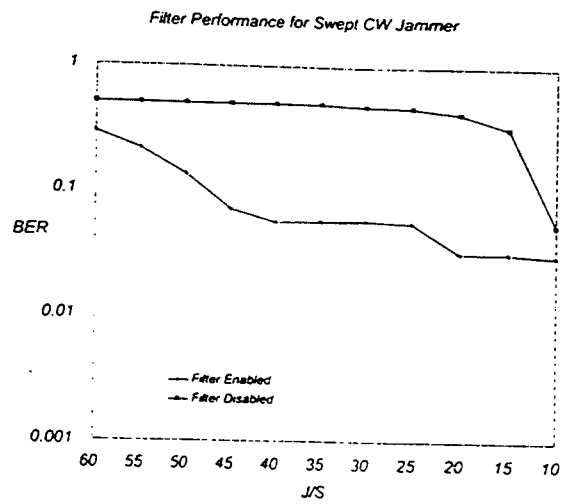


Fig 13 The instantaneous frequency obtained using the zero-crossing technique applied to (a) linear FM, (b) sinusoidal FM and (c) frequency hopping signals, all of 10 dB SNR.

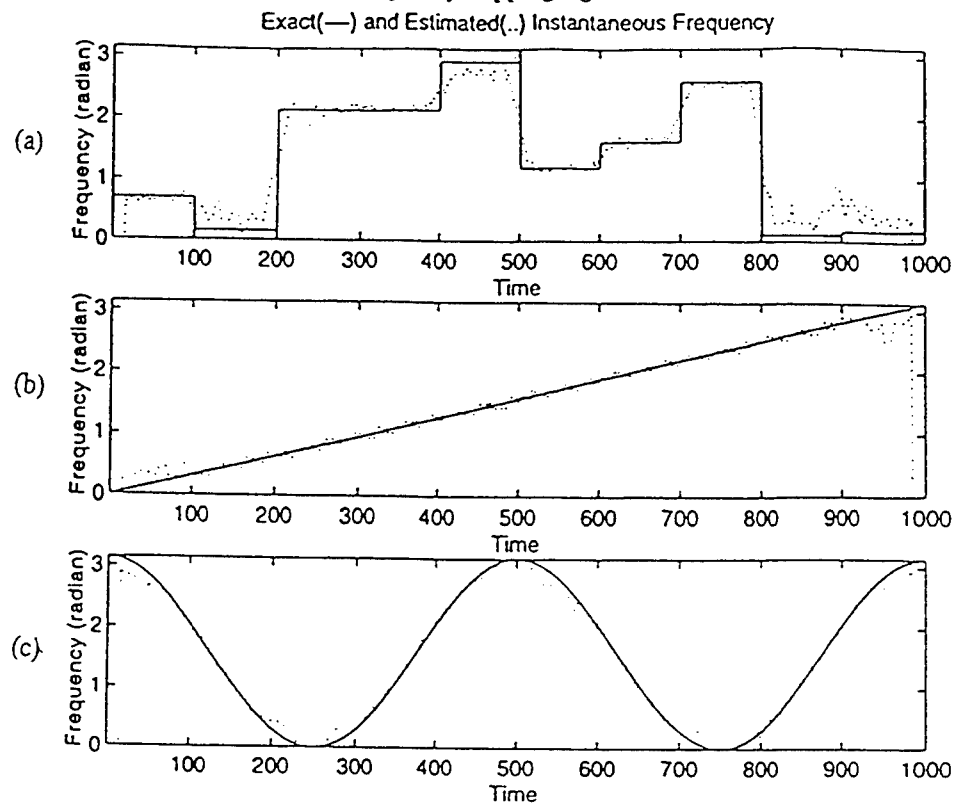
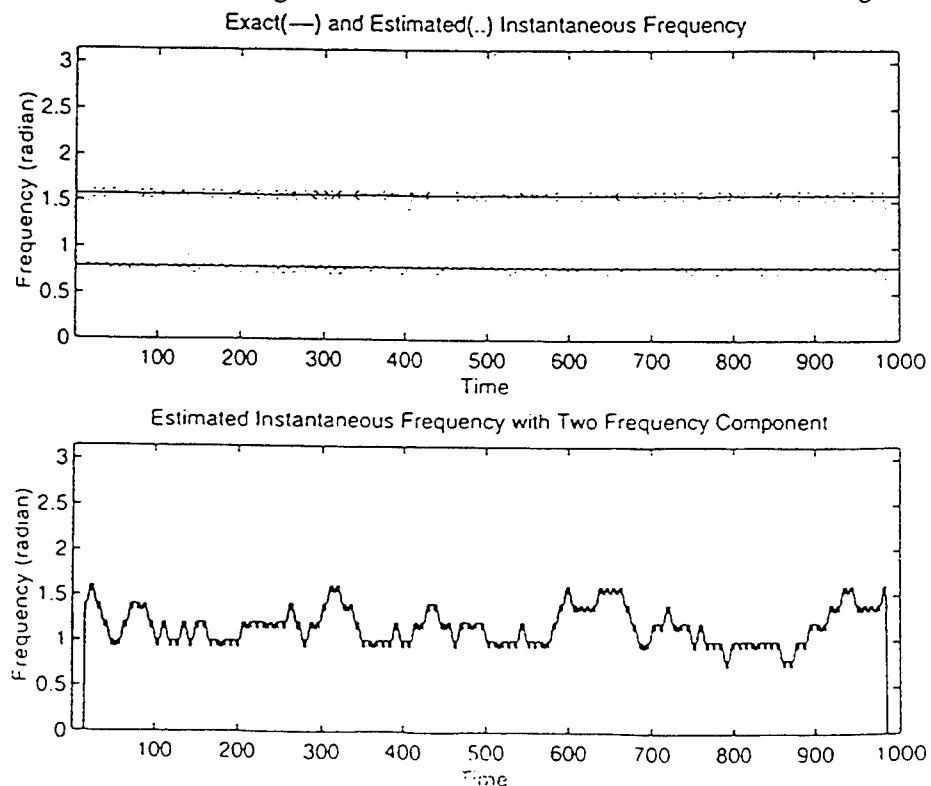


Fig. 14 The zero-crossing estimates of individual and combined sinusoidal signals.



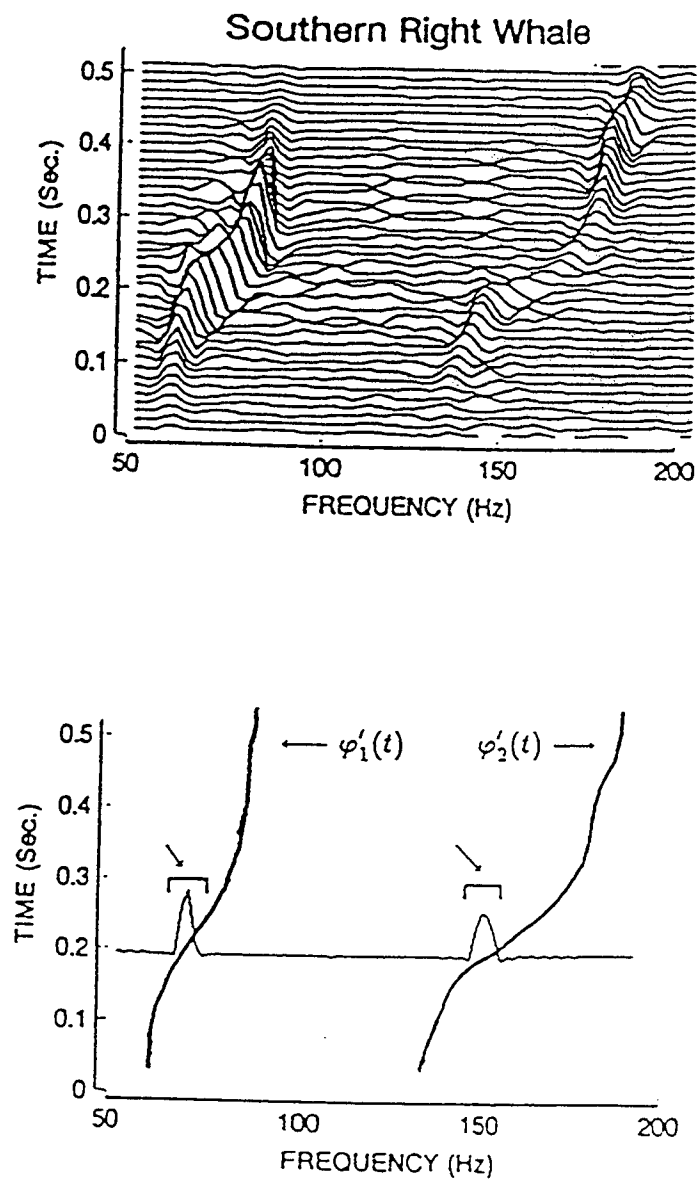


Fig. 15 A multicomponent signal is characterized by the narrowness of both parts in comparison to their separation.

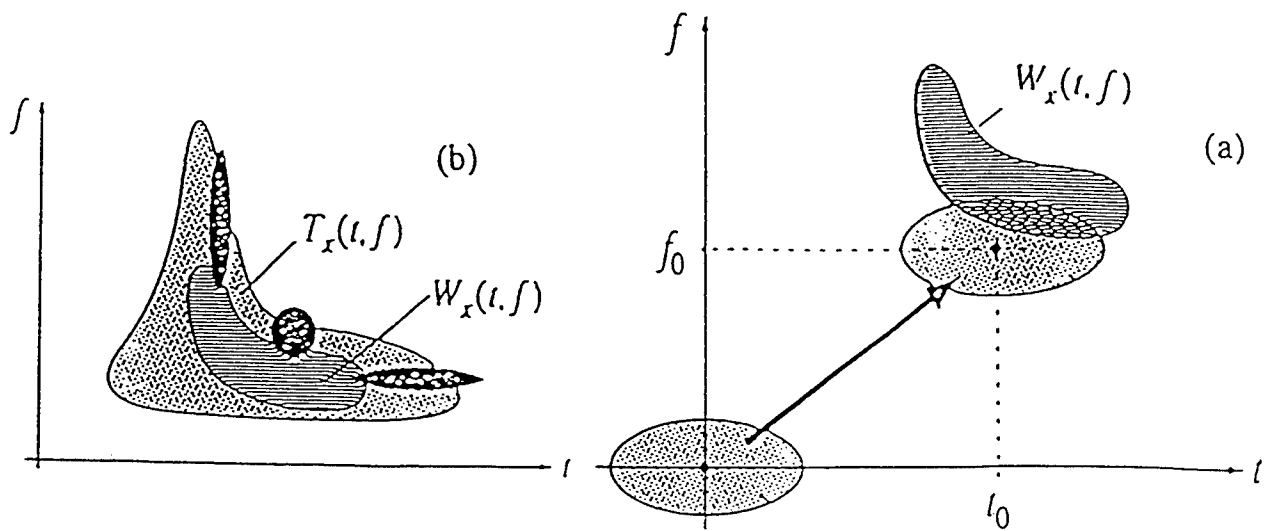


Fig. 16 Quadratic time-frequency distributions (a) Shift-covariance smoothing (b) the affine smoothing.

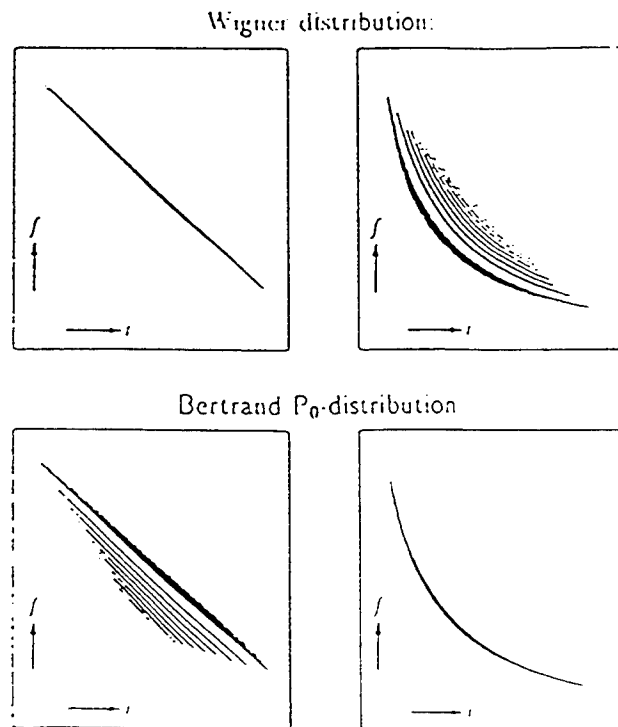


Fig.17 Wigner distribution and the affine smoothing for a chirp and hyperbolic t - f signal geometry.

DESIGNING SOFTWARE BY REFORMULATION USING KIDS

Dr. D. Paul Benjamin
Visiting Assistant Professor
School of Computer and Information Science

Syracuse University
Syracuse, NY 13244

Final Report for:
Summer Research Extension Program
Rome Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base, Washington, D.C.

and

Syracuse University

December 1995

DESIGNING SOFTWARE BY REFORMULATION USING KIDS

D. Paul Benjamin

Visiting Assistant Professor

School of Computer and Information Science
Syracuse University

Abstract

In recent years, much attention has been paid to designing software by formal methods. Such an approach has been shown to facilitate the design of correct software. The KIDS system utilized at Rome Laboratory is an example of a software design system based entirely on a formal approach - the theory of institutions developed by Joseph Goguen and his coworkers. KIDS combines a theory of problem solving with a domain theory to derive a high-level program, which can then be transformed to improve efficiency. Frequently, the problem solving theory is a theory of global search, as it is in the KTS transportation scheduling system. Unfortunately, although global search through the space of possible solutions may be suitable for some problems, it is extremely inefficient for many others. In many problems, the search space is far too large for such an approach. Rather, it is desirable to decompose the space into independent components, so that partial solutions can be found in each piece and combined to give a global solution. KIDS does support problem space decomposition, but does not provide methods for finding useful decompositions for specific problems. This report describes an algebraic approach to theory decomposition and its implementation using Mathematica and GAP. The implementation is illustrated on examples.

D. Paul Benjamin

Introduction

In recent years, much attention has been paid to designing software by formal methods. Such an approach has been shown to facilitate the design of correct software. KIDS (Kestrel Institute Development System) is an example of a software design system based entirely on a formal approach - the theory of institutions developed by Joseph Goguen and his coworkers. Two of the research groups at Rome Labs, the KBSA (Knowledge-Based Software Assistant) group and the KBP (Knowledge-Based Planning) group, use KIDS as the environment within which to develop software.

The emphases of planning and software engineering differ in many ways, e.g., software engineering in a formal environment is concerned with developing a program that is provably correct with respect to a given specification, while planning is concerned more generally with developing a program that has a high probability of producing a specified behavior. However, despite the many different emphases between planning and software engineering, both possess extremely large spaces of possible designs. Thus, designers in both fields are faced with the difficult task of finding efficient designs. A system such as KIDS helps the designer to make clear specifications, to write correct programs, and to refine existing code to improve its efficiency. However, no existing system helps the designer find efficient designs in the first place.

KIDS combines a theory of problem solving with a domain theory to derive a high-level program, which can then be transformed to improve efficiency. Frequently, the problem solving theory is a theory of global search, as it is in the KTS transportation scheduling system. Unfortunately, although global search through the space of possible solutions may be suitable for some problems, it is extremely inefficient for many others. In many problems, the search space is far too large for such an approach. Rather, it is desirable to decompose the space into independent components, so that partial solutions can be found in each piece and combined to give a global solution. KIDS does support problem space decomposition, but does not provide methods for finding useful decompositions for specific problems. This paper describes how a theory of decomposition and reformulation can be used in conjunction with KIDS to derive efficient programs.

Example: n-Queens

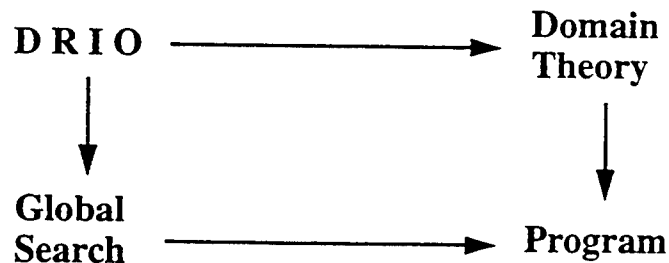
The following brief review of KIDS is intended for the reader with some working knowledge of the

system. Other readers are urged to read Smith(1990) and related papers to understand how KIDS works. This example is drawn from Smith(1990).

The n-Queens task is to place n queens on an nxn chessboard so that no two queens are on the same row, column, or diagonal. We are posed the task of enumerating all solutions for a given value of n.

1		Q		
2				Q
3	Q			
4			Q	
	1	2	3	4

The KIDS user must develop a theory of this domain that specifies the legal solutions. A direct way to do this is to specify a solution as satisfying a predicate that is the conjunction of the goal conditions: no-two-queens-on-same-row & no-two-queens-on-same-column & no-two-queens-on-same-up-diagonal & no-two-queens-on-same-down-diagonal. In Smith (1990), it is shown how KIDS combines this domain theory with a theory of global search to derive a correct high-level program. KIDS provides optimizations that the user can apply to this program. The process of theory combination can be viewed as forming the pushout of the domain theory with the global search theory. DRIO stands for the domain and range of the function to be synthesized, and the input and output conditions that specify legal inputs and outputs.



The goal conditions are used as filters to reduce the search space. However, even with these filters, the search space grows exponentially with the size of the problem. For example, in the 4-queens problem shown above, there is no solution if we place the first queen in the upper left-hand corner or lower-left hand corner, but the system cannot know this until it has searched all ways of placing the remaining queens. This exponential growth in the number of possibilities is the well-known Achilles' heel of problem solving, as it precludes the efficient solution of very large problems. Global search cannot be used to enumerate the solutions to the million-queens problem.

Effective search heuristics are known for finding a single solution to a problem of that magnitude, but they cannot efficiently enumerate all solutions. This prevents consideration of the space of solutions, and in particular prevents optimal problem solving. More seriously, the heuristic approach requires engineering a new heuristic for each problem class. This does not even partially solve the designer's problem of finding an efficient algorithm; it merely transforms it into the problem of finding an efficient heuristic.

What the designer needs is a method for finding good decompositions of the problem. A good decomposition of the problem has at least the following two properties: it splits the problem into components such that solutions within the components compose to form solutions to the whole problem, and it applies to all problems in the class.

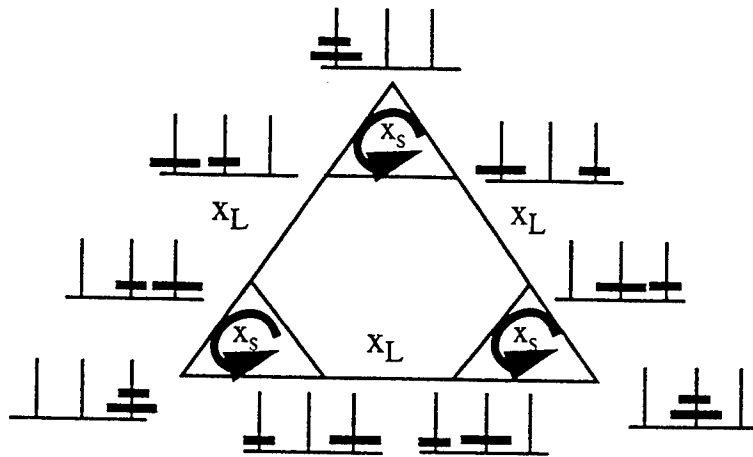
KIDS does support decomposition, with its split and combine operations. However, the only two forms of decomposition implemented are first-rest decomposition and half-half decomposition. In first-rest decomposition, the first part of the problem description is solved and its solution is combined with the solution of the rest (this constituting a recursive call to first-rest decomposition.) In half-half decomposition, the problem description is split roughly in half, and solutions of each half are combined (again leading to recursive calls.) Mergesort is a typical example of half-half decomposition, while bubble sort is an example of first-rest decomposition.

These forms of decomposition are not generally applicable. For example, neither of these forms of decomposition applies to the n -queens problem. In first-rest decomposition, we would place one queen, then place the remaining $n-1$ queens. But as we saw above, if we place the first queen in the wrong place (a corner), the 1-queen solution will not compose with any solution to the $n-1$ queens to give a valid n -queen solution. Similarly, half-half decomposition would find any solution for placing 2 queens, and compose it with another 2-queen solution (which could be a copy of itself.) This need not be a solution to the 4-queen problem.

Formulations of Domain Theories

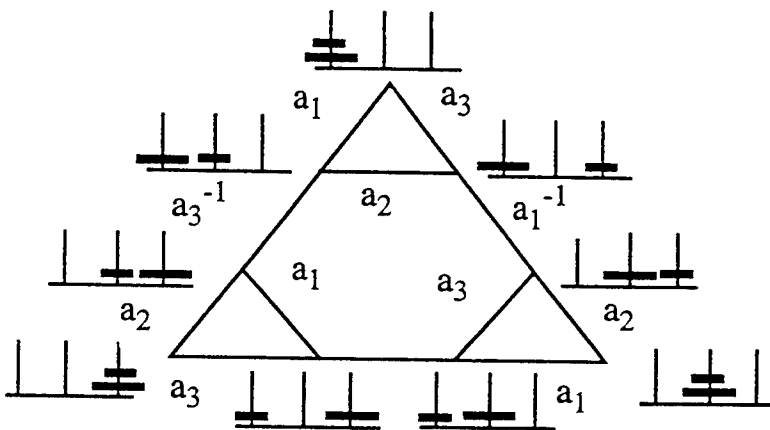
A good choice of notation and a good choice of formulation within that notation are absolutely necessary for effective use of theories. A simple example is given by the following two representations for the

two-disk Towers of Hanoi.



x = Move disk d one peg right
(wrapping around)

x^{-1} = Move disk d one peg left
(wrapping around)



a = Move the top disk on peg p one
peg right (wrapping around)

a^{-1} = Move the top disk one peg
left to peg p (wrapping around)

Even in this simple puzzle there are many possible formulations. The two formulations given above differ in that the first indexes the moves according to the disk moved, whereas the second indexes the moves by the peg moved from (or to, for the inverse moves.) The advantage of the first formulation is clear: it scales up to theories for more disks, because the actions x_S and x_L will be included unchanged in those theories. New moves will be added for the new disks. However, the actions in the second formulation must be redefined as more disks are added, so this theory does not scale up. The result is that analysis and synthesis of the two-disk problem performed using the first formulation can be reused when larger problems are attempted, but analysis and synthesis from the second formulation must be discarded when larger problems are attempted.

Computer science is not the first field to be faced with the problem of properly formulating theories. Throughout the history of science, it has always been desirable to formulate theories in as general a way as possible, so that important regularities are identified and separated from details particular to individual situations. In particular, physics has had a great deal of success in formulating theories of wide generality yet high predictive accuracy. In this paper, we will see that many of the mathematical structures employed in the statement of physical theories can be usefully generalized to the statement of abstract

theories.

We will begin with a brief discussion of the properties of physical theories in the next two sections. The remainder of the paper discusses the appropriate mathematics for analyzing these properties, and gives three simple examples of the analysis and reformulation of theories.

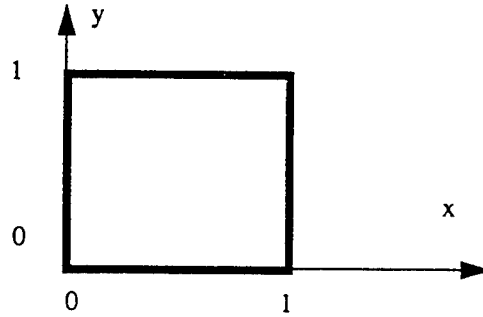
Invariants of Laws

The ability to formulate any law of nature depends on the fact that the predictions given by the law, together with certain initial conditions, will be the same no matter when or where the results of the predictions are observed. In physical theories, the fact that absolute time and location are never relevant is essential for the statement of laws; without this fact, general laws could not be stated, and the complexity of the world would eliminate the possibility of intelligent comprehension of the environment. This irrelevance is stated in terms of the invariance of laws under translation in time and space. Such invariance is so self-evident that it was not even stated clearly until less than a century ago. It was Einstein who recognized the importance of invariance in the formulation of physical law, and brought it to the forefront of physics. Before Einstein, it was natural to first formulate physical law and then derive the laws of invariance. Now, the reverse is true. As the eminent physicist Wigner states, "It is now natural for us to try to derive the laws of nature and to test their validity by means of the laws of invariance. ..."(Wigner, 1967, p.5). This is especially clear in the development of quantum mechanics.

Invariance is important not only in physics. As Dijkstra states, "Since the earliest days of proving the correctness of programs, predicates on the programs's state space have played a central role. This role became essential when non-deterministic systems were considered.... I know of only one satisfactory way of reasoning about such systems: to prove that none of the atomic actions falsifies a special predicate, the so-called 'global invariant'." (Dijkstra, 1985.) In other words, as the system moves in its state space, the global invariant is a law of motion. Dijkstra goes on to point out the central difficulty in the use of invariants: "That solves the problem in principle; in each particular case, however, we have to choose how to write down the global invariant. The choice of notation influences the ease with which we can show that, indeed, none of the atomic actions falsifies the global invariant." We need a mathematics of invariance to help us formulate invariants.

Symmetry

A symmetry is a mapping of an object or system to itself such that the result of the mapping is indistinguishable from the original. For example, the human body (idealized) has a left-right symmetry. The following square has a number of symmetries, including flipping it onto itself about the lines $x = 1/2$ or $y = 1/2$ or $x = y$, and rotating it ninety degrees either clockwise or counterclockwise:



Symmetries can exist in physical space or in state space. For each invariance, there is a corresponding set of symmetries, each of which maintains the invariant. For example, the invariance of physical law under translation in space corresponds to the symmetries of space under all translations. Also, the global invariant of a non-deterministic program corresponds to all permutations of the atomic actions; each permutation maintains the invariant. This correspondence holds in reverse, also. For each set of symmetries, there is a corresponding invariant.

For example, the square above can be represented by the following theory:

$$x = 0, \quad 0 \leq y \leq 1$$

$$x = 1, \quad 0 \leq y \leq 1$$

$$y = 0, \quad 0 \leq x \leq 1$$

$$y = 1, \quad 0 \leq x \leq 1$$

This theory has syntactic symmetries corresponding to the symmetries of the square, e.g. interchanging x and y gives the flip about the line $x = y$, and interchanging the first and second lines of the theory flips the square about the line $x = 1/2$.

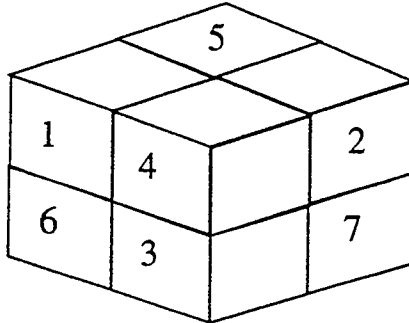
Many of the important symmetries in physics are geometric. In other words, they are symmetries of the space in which motion takes place. By viewing a program as “moving” in its state space, we can take the same approach as physicists: formulate geometric symmetries of the space, and use them to derive invariants, thereby obtaining laws governing the use of the program.

The Mathematics of Symmetry: Groups Theory

Given a global invariant, the corresponding set of transformations is closed under composition, and for every transformation, its inverse is also a transformation that preserves the invariant. The identity transformation is also always in the set. Thus, the set of transformations form a *group*. Group theory is the language of symmetries, and has assumed a central role in modern physics. Group theory can also be used to analyze the symmetries of a task and derive invariants, which are then used to synthesize a program. The following example is given in Benjamin (1994). (This paper will not provide any background

in group theory. The reader is referred to any standard text.)

Let us begin by examining a simple task, the 2x2x2 Rubik's Cube with 180-degree twists (we use such a small example for brevity of presentation, but the techniques are generally applicable, as will be shown). Let the 8 cubicles (the fixed positions) in the 2x2x2 Cube be numbered in the following way (8 is the number of the hidden cubicle):



The goal configuration for the 2x2x2 Rubik's Cube with 180-degree twists.

Number the cubies (the movable, colored cubes) similarly, and let the goal be to get each cubie in the cubicle with the same number. For brevity of presentation, we will consider only 180° twists of the cube. Let f , r , and t denote 180° clockwise turns of the front, right, and top, respectively (cubie 8 is held fixed; Dorst (1989) shows that this is equivalent to factoring by the Euclidean group in three dimensions). Note that this cubie numbering is just a shorthand for labeling each cubie by its unique coloring. This holds true for the Cube with only 180° twists, as position determines orientation.

The actions for the Cube can be represented as a group, which is generated by the actions f , r , and t . We use group representation theory to represent f , r , and t as matrices:

$$f = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad r = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix} \quad t = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

These matrices are 7-dimensional, corresponding to the 7 unsolved cubies. We find eigenvectors of eigenvalue 1; these are the invariants. Any invariant of all the actions is irrelevant and can be projected away. To do this, we first change the coordinate system so that the invariant eigenvectors are axes, and then project to the noninvariant subspace, removing all irrelevant information at once. In this case, the eigenvectors are:

$$r: \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \text{ for } \lambda = 1, \text{ and } \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix} \text{ for } \lambda = -1 \quad f: \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \text{ for } \lambda = 1, \text{ and } \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \text{ for } \lambda = -1$$

$$t: \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \text{ for } \lambda = 1, \text{ and } \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} \text{ for } \lambda = -1 \quad \text{and the common invariant eigenvector is: } \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

Note that we have shortened these eigenvectors to save space; they are actually 7-vectors, with additional zeroes. We then change the basis. The appropriate matrix is:

$$P = \begin{bmatrix} \frac{1}{\sqrt{3}} & 0 & 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & 0 & 0 & 0 \\ \frac{1}{\sqrt{3}} & 0 & 0 & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & 0 & 0 & 0 \\ \frac{1}{\sqrt{3}} & 0 & 0 & 0 & \frac{2}{\sqrt{6}} & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & 0 & -\frac{1}{2} & -\frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & 0 & -\frac{1}{2} & \frac{1}{2} & -\frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & 0 & \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} & 0 \end{bmatrix}$$

yielding the new representations for r, f, and t:

$$r = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{\sqrt{3}}{2} & 0 & 0 & 0 \\ 0 & 0 & \frac{\sqrt{3}}{2} & -\frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 \end{bmatrix} \quad f = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & -\frac{\sqrt{3}}{2} & 0 & 0 & 0 \\ 0 & 0 & -\frac{\sqrt{3}}{2} & -\frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 \end{bmatrix} \quad t = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

This procedure computes the irreducible invariants of a group. The irreducible factors of dimension 1, 1, 2, and 3 are found along the diagonals of the matrices. Projecting to these subspaces yields two subproblems:

$$r = \begin{bmatrix} \frac{1}{2} & \frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & -\frac{1}{2} \end{bmatrix} \quad f = \begin{bmatrix} \frac{1}{2} & -\frac{\sqrt{3}}{2} \\ -\frac{\sqrt{3}}{2} & -\frac{1}{2} \end{bmatrix} \quad t = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \quad r = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & -1 & 0 \end{bmatrix} \quad f = \begin{bmatrix} 0 & 0 & -1 \\ 0 & 1 & 0 \\ -1 & 0 & 0 \end{bmatrix} \quad t = \begin{bmatrix} 0 & -1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

On cubelets 1, 2, and 3, the subgroup generated is
{i, r, f, t, rt, tr}

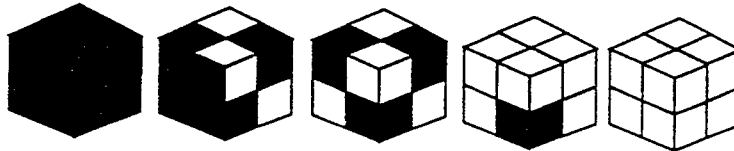
On cubelets 4, 5, 6, and 7, the subgroup generated is
{i, r, f, t, rf, rt, fr, ft, tr, tf, rfr, rft, rtr, rtf, frt, ftr, ftf, trf, tfr, rfrt, rftt, rtrf, rtrr}

Using each set of matrices as generators, we get two subgroups of actions, the second of which is a faithful representation of the whole group. The first subgroup moves cubies 1, 2, and 3, while holding 4, 5, 6, and 7 in position. The second subgroup moves cubies 4, 5, 6, and 7 while holding 1, 2, and 3 in their positions. We then repeat this procedure on the first set of actions to obtain a full set of prime factors of the group.

These factors can be assembled in different ways to form serial algorithms. There is more than one way to decompose this group, analogous to the different ways of listing the prime factors of a number. Five serial algorithms are obtained in this way. We examine two of them.

$$R = \{i, frtr\} \circ \{i, rtft\} \circ \{i, t\} \circ \{i, r, f\}$$

Serial Algorithm 1:



One of $\{i, r, f\}$ brings cubelet 3 into cubicle 3.

One of $\{i, t\}$ brings cubelet 1 and 2 into their places.

One of $\{i, rtft\}$ brings (4,6) and (5,7) in the proper planes ('the front face looks right').

One of $\{i, frtr\}$ finishes the Cube.

The above figure is read right-to-left; shaded cubicles have been solved. "i" denotes the identity (null) action. The average number of moves required to solve the Cube in this way is 5.17.

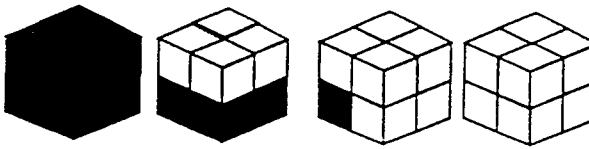
We see that there are two "subspaces" of cubies in this Cube: one consisting of cubicles 1,2,3 and 8, and the other of cubicles 4,5,6, and 7. For each subspace, there is a subprogram that interchanges the cubies in its cubicles while holding the cubies in the other subspace invariant. We are thus justified in thinking of these two sets of cubicles as *independent* subspaces. Recognizing the symmetries that characterize such subspaces is essential for synthesizing such algorithms.

Each step in the algorithm brings a subset of cubies to its goal value. Subsequent steps hold that pattern of cubies invariant. In this way, a divide-and-conquer algorithm is synthesized. For example, the first step solves cubicle 3. Knowing the colors of the solved cubicle 8, we know the colors of cubicle 3 - it has the same color as the bottom of cubicle 8, and two new colors. There is only one such cubie, and it must be in one of three locations: in its goal position, or in cubicle 1 or 2. The system need only examine those 3 values to determine what action to take. Once cubicle 3 is solved, it need not be examined again. The system next solves cubicles 1 and 2; it need only examine either position to see if the proper cubie is there; if so, it does nothing, otherwise, it twists the top. Finally, the system uses the appropriate sequence of actions to solve the remaining four cubicles, by first examining the front face to see if it is a uniform color, and then examining the top or right face to see if it is of uniform color.

We have transformed the global 7-dimensional description of the Cube into a composition of local descriptions, each characterized by a set of symmetries. This decomposition has average cost of 5.17, whereas an optimal solution is of average length 2.46. There are better decompositions. We now examine the best decomposition.:

$$R = \{i, t\} \circ \{i, r, rt\} \circ \{i, f, ft, fr\}$$

Serial Algorithm 2:



One of $\{i, f, fr, ft\}$ brings cubelet 6 into cubicle 6.

One of $\{i, r, rt\}$ brings 3,7 in place (bottom layer correct).

One of $\{i, t\}$ finishes the Cube.

The average number of moves to solve the Cube using this decomposition is 2.75.

Each decomposition can be thought of as a coordinate system whose origin is the goal state. For example, the second serial algorithm can be thought of as a 3-dimensional coordinate system (a,b,c) where a is in $\{i,f,ft,fr\}$, b is in $\{i,r,rt\}$, and c is in $\{i,t\}$.

From the Cube example, we see that we can view a task representation as a coordinate system whose axes are the components of the task. Using group representation theory, we represent the actions as matrices. Changing the basis so that invariant eigenvectors are axes re-expresses the task space so that subtasks are coordinate subspaces, thereby identifying a good decomposition. The divide-and-conquer algorithm transforms the group of the problem into a product of smaller groups (a composition series). This is done by factoring the group at each step by an irreducible normal subgroup, yielding a quotient group. This procedure is similar to that used in Galois theory, in which an equation is solvable by extraction of roots iff its group of symmetries is solvable. However, the situation with domain theories is more general, in that a divide-and-conquer algorithm can be generated even if the transformation group is not solvable, because all that we require is that the decomposition yield components that are small enough that global search can be applied efficiently to each of them.

The Mathematics of Local Symmetries: Inverse Semigroups

The previous section illustrates how the analysis of global symmetries can be useful in synthesis. However, global symmetries, and in general global properties, are not sufficiently general. We must also examine *local symmetries*, which are symmetries between parts of a system, rather than of the whole system. Local symmetries correspond to local invariants, which are predicates that hold for some part of a system, but not necessarily for the whole system. An example is a loop invariant in a program, which

holds during the execution of the loop, but not necessarily elsewhere. This is the sort of invariant more often encountered in computing.

To handle local symmetries, we use a more general theory than group theory: the theory of inverse semigroups. A semigroup is a closed, associative system, and an inverse semigroup has the additional property that every action is invertible. The difference between a group and an inverse semigroup is that the transformations in the group must be globally defined (they correspond to global symmetries) whereas those in the inverse semigroup can be defined on only part of the space, and are thus partial functions on the space. Inverse semigroups are thus more appropriate for reasoning about programming constructs, e.g., rules, which can be defined only on some variable bindings.

In physical theories, space-time is represented in terms of a coordinate system. Invariance under translation in time or space then becomes invariance under coordinate change. Inverse semigroups possess a similar notion of coordinate system, and we use this in the same way.

We consider a task theory $M = (Q, A, \delta)$ consisting of a set A of actions defined as partial functions on a state space Q , together with a mapping $\delta: Q \times A \rightarrow Q$ that defines the state transitions. We are not concerned with the syntactic details of the encoding of the actions A , but rather with which actions should be labeled the same and should therefore be considered instances of a common abstraction. In other words, we are concerned with the algebraic structure of A . A task is specified by a pair (i, g) , where $i: 1 \rightarrow Q$ maps a one-element set into Q identifying an initial state, and $g: 1 \rightarrow Q$ identifies a desired state. Without loss of generality, we can restrict our attention to semigroups of partial 1-1 functions on the state space Q (Howie, 1976). This formalism is extremely general. It encompasses nondeterministic systems and concurrent systems.

The description of a system for the synthesis of a plan (or control) for M differs from the description of M in at least one essential way: the process of planning an action is reversible whether or not the action itself is reversible (assuming the synthesizing system can backtrack.) Thus, the process of synthesizing plans can be described by a theory whose actions form an appropriate inverse semigroup containing the original actions together with newly added inverses corresponding to backtracking. As this paper is primarily concerned with system synthesis, we will analyze only inverse semigroups in the remainder of this paper.

To analyze the structure of such a semigroup of transformations, a usual step is to examine Green's relations (Lallement, 1979). Green's equivalence relations are defined as follows for any semigroup S :

$$\begin{aligned} a R b &\text{ iff } aS^1 = bS^1 & a L b &\text{ iff } S^1a = S^1b & a J b &\text{ iff } S^1aS^1 = S^1bS^1 \\ H &= R \cap L & D &= RL \end{aligned}$$

where S^1 denotes the monoid corresponding to S (S with an identity element 1 adjoined). Intuitively,

we can think of these relations in the following way: aRb iff for any plan that begins with “a”, there exists a plan beginning with “b” that yields the same behavior; aLb iff for any plan that ends with “a”, there exists a plan ending with “b” that yields the same behavior; aHb indicates functional equivalence, in the sense that for any plan containing an “a” there is a plan containing “b” that yields the same behavior; two elements in different D-classes are functionally dissimilar, in that no plan containing either can exhibit the same behavior as any plan containing the other.

Green’s relations organize the actions of a transformation semigroup according to their functional properties, and organize the states according to the behaviors that can be exhibited from them. This allows us to define a basic local neighborhood of a semigroup:

Definition. Given a transformation semigroup $S = (Q, A)$ and a point $q \in Q$, a D-class of S containing an action whose domain includes q is called a *basic local neighborhood* of q .

Each basic local neighborhood in state space consists of behaviorally similar states. If the agent’s perceptual capabilities are not sufficiently precise to distinguish different neighborhoods, then it cannot plan and move effectively in the space. A state q may be in more than one basic local neighborhood. Clearly, every element of Q is in at least one basic local neighborhood. The set of all basic local neighborhoods forms a neighborhood base for a topology on Q ; however, this direction will not be pursued in this paper, so we will often just refer to basic local neighborhoods as neighborhoods.

Utilizing Green’s relations, Lallement defines a local coordinate system for a neighborhood:

Definition. Let D be a D-class of a semigroup, and let $H_{\lambda\rho}$ ($\lambda \in \Lambda, \rho \in P$) be the set of H-classes contained in D (indexed by their L-class and R-class). A coordinate system for D is a selection of a particular H-class H_0 contained in D , and of elements $q_{\lambda\rho}, q'_{\lambda\rho}, r_{\lambda\rho}, r'_{\lambda\rho} \in S^1$ with $\lambda \in \Lambda, \rho \in P$ such that the mappings $x \rightarrow q_{\lambda\rho}xr_{\lambda\rho}$ and $y \rightarrow q'_{\lambda\rho}yr'_{\lambda\rho}$ are bijections from H_0 to $H_{\lambda\rho}$ and from $H_{\lambda\rho}$ to H_0 , respectively. A coordinate system for D is denoted by $[H_0; \{(q_{\lambda\rho}, q'_{\lambda\rho}, r_{\lambda\rho}, r'_{\lambda\rho}) : \lambda \in \Lambda, \rho \in P\}]$.

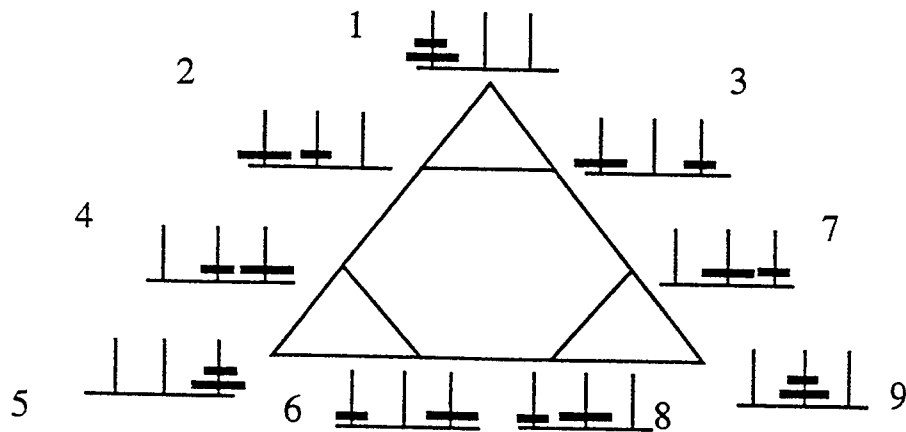
There may be more than one local coordinate system for a D-class. Each coordinate system gives a matrix representation in much the same way that a coordinate system in a vector space gives a matrix representation, permitting us to change coordinates within a local neighborhood by performing a similarity transformation (inner automorphism) in the usual way (the reader is referred to Lallement for details).

Each local coordinate system within the semigroup expresses a distinct syntactic labeling of a subsystem, as we can choose a point in the neighborhood to be the “origin”, and label points in the neighborhood according to the actions that map the origin to them. In addition, each coordinate system can be used to create a matrix representation for the semigroup, in much the same way that a coordinate system for a vector space yields matrices for the linear transformations of the vector space. The reader is

referred to Lallement (1974) for details.

Example: The Towers of Hanoi

Let us number the nine states of the 2-disk Towers of Hanoi as follows:



Let the two possible actions be denoted by "x" and "y".

x = 1 2 3 4 5 6 7 8 9 y = 1 2 3 4 5 6 7 8 9
 2 3 1 5 6 4 8 9 7 4 8 3

"x" moves the small disk right one peg (wrapping around from peg 3 to peg 1), and "y" moves the large disk one peg to the left (wrapping around from peg 1 to peg 3). In the figure, "x" is shown by narrow, counterclockwise arrows, and "y" is shown by thick, counterclockwise arrows. These two actions generate a semigroup of 31 distinct partial functions on the states. Green's relations for this semigroup are:

D0			0
D1			x, xx, xxx
D2	xyx xyxxyx xyxxyxxyx	xy xyxxy xyxxyxxy	xyxx xyxxyxx xyxxyxxyxx
	xxyx xxyxxyx xxyxxyxxyx	xxxy xxxyxxy xxxyxxyxxy	xxxyxx xxxyxxyxx xxxyxxyxxyxx
	yx yxxyx yxxyxxyx	y yxxy yxxyxxy	yxx yxxyxx yxxyxxyxx

There are three D classes, shown as the three separate large boxes. In each D class, the R classes are rows and the L classes are columns, and they intersect in the small boxes, which are H classes. Note that D0 and D1 consist of only one R class and one L class, and hence one H class. The idempotents are in bold type.

There are no nontrivial inner automorphisms of D0 and D1. The group of inner automorphisms of D2 is a cyclic group of order three. These coordinate transformations are global within D2, but are local in the semigroup. These inner automorphisms are calculated by the matrix techniques explained by Lallement (1979). A generator for this group is the automorphism that maps xyx to xyx , xyx to yxx , and yxx to xyx . Factoring D2 by this map gives:

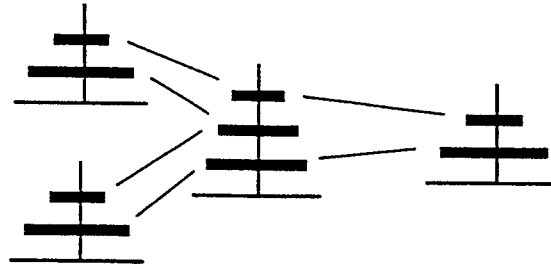
Define $z = \text{case } \{ \text{little disk left of large disk: } xyx$
 little disk on large disk: xyx
 little disk right of large disk: $yxx \}$

where z is a new symbol.

x is as before, and z moves both disks left one peg. z is a macro-action, which is implemented as a disjunction of sequences of actions. x and z are independent controls; x solves the position of the small disk, and z solves the big disk. x does not change the position of the big disk, and z does not change the relative positions of the disks. The disks can be solved in either order, as the new representation is an abelian group. This new representation captures the important property of the Towers of Hanoi task: the disks can be solved in any order. In the original theory, this property was obscured by details of the implementations of the disk moves.

This decomposition applies not only to the two-disk problem, but to all Towers of Hanoi problems. This is seen by forming free products of the semigroup with itself, amalgamating the coordinate axes in all possible ways. There are three free products of this semigroup with itself that amalgamate coordinates: D2 can be identified with itself, D1 with itself, and D2 with D1 (by mapping xyx , xyx , and yxx to x). These correspond to identifying the moves of the larger disk of one copy of the semigroup with the moves of the larger disk in another copy, identifying the moves of the smaller disk with the moves of the smaller disk, and identifying the moves of the larger disk with the moves of the smaller disk, respectively. The result of this construction is the three-disk Towers of Hanoi, which is viewed as three copies

of the two-disk task running concurrently.





From the logical perspective, this can be viewed as composing three copies of a theory for the two-disk Towers of Hanoi, to yield a valid theory for the three-disk problem. Copies of the theory are joined by unifying variables between theories. Two copies joined at the middle disk will not suffice, as then the largest disk could be placed on the smallest disk. A third copy of the theory prohibits this. The purpose of computing a coordinate system is that the coordinate axes determine what to unify.

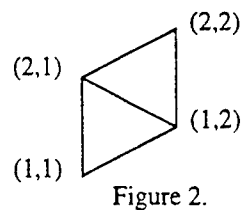
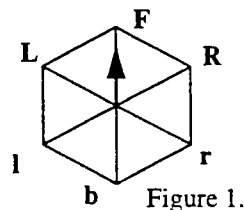
The interaction of these three smaller tasks yields a set of relations. For example, consider moving the middle disk one peg to the right. When considering this disk as the larger disk in the task consisting of the upper two disks, this move is $xyxy$ (in the state when all disks are on the left peg), $yxyx$ (when the smallest disk is to the right of the middle disk), or $xyxyxyx$. On the other hand, when considering this disk as the smaller disk in the task consisting of the lower two disk, this move is x (in the other copy of the semigroup). This means that we add the relation $xyxy = yxyx = xyxyxyx$ to the presentation of the semigroup.

This method of composing larger tasks from smaller ones guarantees that this decomposition generalizes to any Towers of Hanoi problem with n disks, by considering a coordinate system for the three-disk task, and forming free products with amalgamation in the same manner as before. These free products with amalgamation will always reduce to free products with amalgamation of coordinate systems of the two-disk task, so that by induction the properties of the two-disk task determine the properties of all Towers of Hanoi tasks.

Example: Robot Path Planning

Let us consider a simple robot that moves on a hexagonal grid. Each grid point is labeled by its row and column. The robot's orientation can be north (N), northeast (NE), southeast (SE), south (S), southwest (SW), or northwest (NW). It has six basic motions: forward (F), forward right (R), forward left (L),

Figure 1.  Figure 2. 



0					
FbRr	Rb	Fr			
FrF	FrR	FbR			
bRb	bFr	bRr			
bFrR	Rb	rF			

FRrb	RIFL	FRRF	FRr	RbLI	RIF	RbL	FR	RI	RbLb	RIL	FRR
rrb	FrLbl	rRF	Fbrr	FrLI	FrFIF	FrL	Fbr	FrFI	FrLb	rrbR	FbrR
FILr	Lbl	LIR	Rbrr	RrLI	LbII	RrL	Rbr	FrRI	RrLb	FrRIL	RbrR
lbLr	LFbl	LFIR	lbLrF	LFI	lbF	LF	LFIRbr	lb	LFb	lbL	LFIRb
brrb	rLbl	brRF	brr	rLI	rFIF	rL	br	rFI	rLb	rFIL	brR
ILr	IFL	bLIR	rFRr	bRrLI	rRIF	bRrL	rFR	rRI	IFLL	rRIL	rFRR
Rrb	bRIFL	RRF	bFRr	RRFr	bRIF	bRbL	bFR	bRI	bRbLb	bRIL	bFRR
bLr	bFL	bLbR	bLrF	LLFI	bFLI	LLF	bLrFR	Llb	bFLL	bLrR	bLbRb
IFr	bl	IR	IFrF	IRr	bII	bILF	IRbr	bIIb	bIL	IFrR	IRb
rb	RFrLbl	RF	rbF	RFr	rbRIF	RFrL	RFbr	rbRI	RFrLb	rbR	RFb
Lr	FL	LbR	LrF	LbRr	FLI	FLLF	LrFR	FLlb	FLL	LrR	LbRb
FIFr	Fbl	FIR	FIFrF	FIRr	FbII	ILF	FIRbr	IIb	FbIL	IIbL	FIRb

Green's relations are used to construct a Rees matrix representation for the semigroup, and for the global and local inner automorphisms (the coordinate transformations.) These details are omitted here, due primarily to space considerations. Factoring this semigroup by its local coordinate transformations gives the familiar decomposition of path planning into planning changes in position and then planning changes in orientation. (More precisely, the semigroup is the normal extension of the subsemigroup that holds orientation invariant by a group of orientation changes that holds position invariant.)

Now let us see how this decomposition scales up to all path planning on any such grid.

There are twelve free products of this semigroup with itself that amalgamate local coordinates.

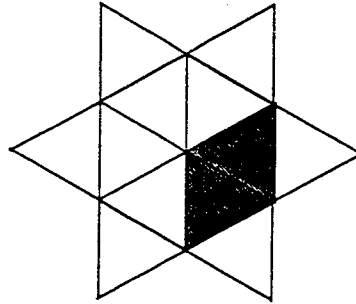


Figure 3. The twelve ways of composing 2x2 grids. This "flower" shape contains twelve 2x2 grids, each of which intersects every other in at least one point, the center point.

For example, the state (2,1,N) of the bold grid in Figure 3a corresponds to a different state in each of the other eleven 2x2 grids in the figure. For example, it corresponds to the state (2,1,SE) in the bold 2x2 grid in Figure 3b. In this way, each state in the 2x2 grid can potentially correspond to any of the other eleven states in another 2x2 grid. For example, the motion $RRFrLbll$, which maps (2,1,N) to (2,1,SW) must be given the same relabeling as $LbllbLrF$, which maps (2,1,SE) to (2,1,N). Continuing in this way, we derive the following: $FRRF = rLbL = lLr = RFrL = bllb = LrFR = RRFrLbll = bLrFRRfr = LbllbLrF = FrLbllbL = rFRRFrLb = llbLrFRR$. These twelve strings are the motions that rotate the robot counter-clockwise one turn, while holding the grid position invariant. Similarly, the clockwise motions, and the motions that hold the orientation invariant while changing position map together.

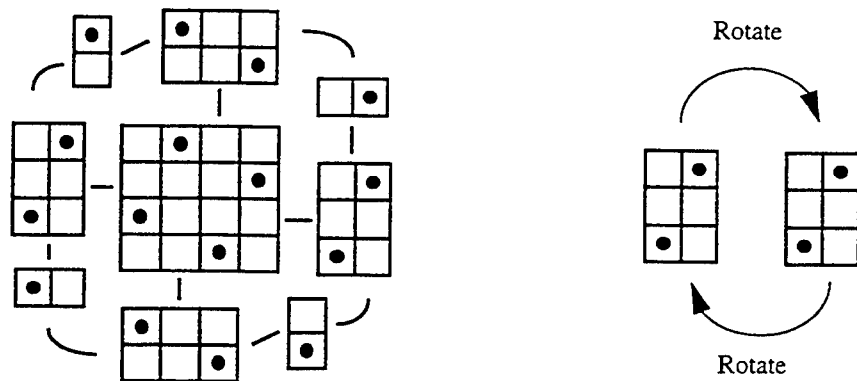
This can be viewed as composing twelve copies of a theory for the 2x2 grid to yield a valid theory for moving on a larger grid. Copies of the theory are joined by unifying variables between theories. The purpose of computing a coordinate system is that the coordinate axes identify what to unify. When the small grid is a subtask of a larger grid, the actions on the small grid must be able to be formulated as actions on the large grid. So, when considering all ways of composing grids to make larger grids, we are considering all ways of formulating small grid actions as larger grid actions, i.e., all ways of transforming coordinates within the small grid.

Thus, this method of composing larger tasks from smaller ones guarantees that the abstract properties of the small grid scale up to any grid composed from small grids, and in particular that the decomposition of the small grid generalizes to this infinite class of grids. This justifies the use of this path planning algorithm on any such grid. If any of the component semigroups resulting from this reformulation are too large, then the method is applied recursively, yielding a hierarchical decomposition. The method is terminated when the semigroups at the leaf nodes are considered sufficiently small to be searched, or are recognized as previously solved subtasks.

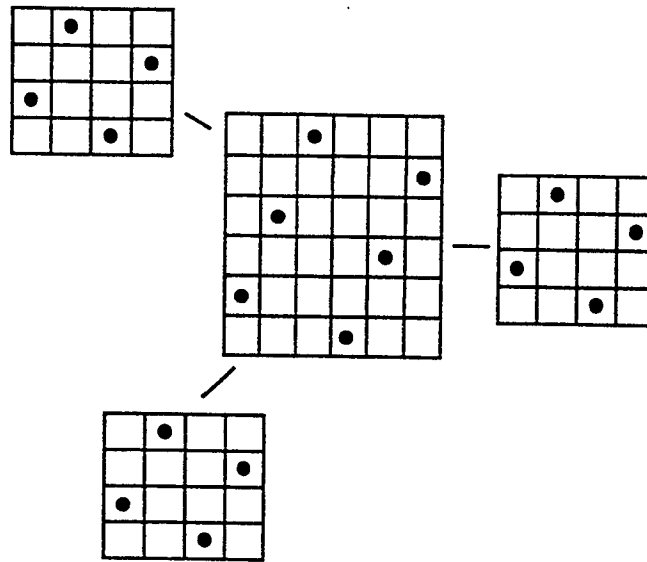
Example: N-Queens

We now return to the n-Queens problem, to show how the reformulation method leads to a program for enumerating the solutions to the problem without considering any non-solutions. In this section, we will not show the computation of Green's relations, as this is repetitive. Instead, we will show the local and global symmetries pictorially.

The next figure shows a solution to the 4-queens problem. (The other solution is obtained by a flip symmetry of this one.) This solution has a hierarchy of local symmetries. The top-level local symmetry is that of the 3x4 subproblems, but the four 3x4 subproblems in the 4x4 are mapped to each other only by the cyclic global symmetry of order 4. The first new local symmetries arise from considering the 2x3 subproblem that contains 2 queens. The four instances of this subproblem are of course also mapped to each other by the cyclic global symmetry, but also map to themselves by a local symmetry of order 2. The next figure shows these four instances, together with the 1x2 subproblems that arise from intersecting the 2x3 subproblems:



As in the other problems considered, we construct larger instances of the n-queens problem from copies of the 4x4 solution, amalgamating the local symmetries. For example, the 6-queens solution can be constructed from four copies of the 4-queens solution by amalgamating 2x3 subproblems:



The 6-queens solution can be viewed as three 4-queens solutions in this way.

Stated another way, the construction of the 6-queens solution from multiple 4-queens solutions can be thought of as follows: take 2 copies of the 4-queens solution and amalgamate a 2x3 subproblem in one copy with a 2x3 in the other. There are 6 distinct queens. The four queens in the first copy are guaranteed to be on their own rows, columns, and diagonals, and the four queens in the second copy are similarly guaranteed to satisfy the solution conditions. But we are not guaranteed that the two non-amalgamated queens in one copy are on distinct rows, columns, or diagonals from the two non-amalgamated queens in the other copy. It is necessary that those four queens also satisfy the solution conditions, i.e., we need them to be a 4-queens solution by themselves. Therefore, composing three 4-queens solutions in this way is both necessary and sufficient to give a 6-queens solution, which implies that all 6-queens solutions are constructed in this way.

Thus, we can construct n -queens solutions from smaller solutions. It is not necessary to search a space containing non-solutions. Instead, analysis of the structure of the base case reveals both the decompositions of the base case and the inductive definition of the class of problems with the same structure. Ideally, a system such as KIDS should have the capability to perform this analysis, to synthesize programs that construct the solution space directly in this fashion whenever possible.

Implementation

This reformulation procedure has been implemented using the Mathematica system coupled with the GAP system (Groups, Algorithms, and Programs). This has permitted a number of examples to be computed, including those in this report, demonstrating the feasibility of this approach.

Two Mathematica packages were utilized in the development: the Combinatorica package, which provided routines for manipulating graphs and their symmetries, and the Automata package, which provided routines for computing state trajectories and transformation semigroups, as well as some code for Green's relations. Custom code was written for computing Green's relations in a more general setting, and for manipulating presentations of semigroups.

The GAP system provides routines for group theory as well as a large library of finite groups, which were utilized in the analysis of symmetries. GAP is called through MathLink, which is the backend of Mathematica. Mathematica serves as the frontend because of its fine graphics.

All the development was done on the KBSA computers at the Rome Laboratories.

Technical Summary

A good formulation of a domain theory omits irrelevant detail and identifies the essential structure of the domain. In particular, it identifies the subgoal decompositions that lead to efficient solutions. This paper has described a research project exploring the mathematical foundations of theory formulations. An analogy is constructed between state spaces and physical space. In this analogy, a domain theory gives the laws of motion in the state space, and a coordinate system gives a formulation of a theory. The formulation of theories is then governed by the established principles of symmetry and invariance. The symmetries of state spaces of base cases can be used to formulate invariants for recursive programs that solve an entire class of tasks. The goal of this research is a comprehensive theory of the symmetries and invariants of programs, and an implemented system for program synthesis based on the theory.

Part of this theory has been implemented in Mathematica and GAP, and applied to domain theories of the type used in the KIDS software synthesis system. KIDS supports the divide-and-conquer approach to problem solving, but needs a technique for automatic identification of useful decompositions. This work has completed a theory of decomposition, and demonstrated how it can be used to reformulate KIDS domain theories to find effective algorithms.

References

- Benjamin, D. Paul, (1994). Formulating Patterns in Problem Solving, *Annals of Mathematics and Artificial Intelligence*, **10**, pp.1-23.
- Benjamin, D. Paul, (1992a). Towards an Effective Theory of Reformulation, in *Proceedings of the Workshop on Change of Representation and Problem Reformulation*, Michael R. Lowry (ed.), NASA Ames Research Center Technical Report FIA-92-06, pp.13-27, April, 1992.
- Benjamin, D. Paul, (1992). Reformulating Path Planning Problems by Task-preserving Abstraction, *Journal of Robotics and Autonomous Systems*, **9**, pp. 1-9.
- Dijkstra, E. W., (1985). Invariance and non-determinacy, in *Mathematical Logic and Programming Languages*, Hoare and Shepherdson, eds., Prentice-Hall, pp.157-165.

- Dorst, Leo, (1989). Representations and Algorithms for the 2x2x2 Rubik's Cube, Philips Technical Report TR-89-041.
- Howie, J. M., (1976). An Introduction to Semigroup Theory, Academic Press.
- Lallement, Gerard (1979). Semigroups and Combinatorial Applications, Wiley & Sons.
- Petrich, Mario, (1984). Inverse Semigroups, John Wiley & Sons, Inc., New York.
- Smith, Douglas R., (1990). KIDS: A Semi-Automatic Program Development System, *IEEE Transactions on Software Engineering*, Vol. 16, No. 9, Special Issue on Formal Methods. pp.1024-1043, September, 1990.
- Wigner, P. (1967) Symmetries and Reflections, Indiana University Press.

Ajit Choudhury report unavailable at time of publication.

Computer-Aided-Design Program for Solderless Coupling
Between Microstrip and Stripline Structures

Frances J. Harackiewicz
Associate Professor
Department of Electrical Engineering

Daniel K. Lee
Ph.D. candidate

Byungje Lee
Ph.D candidate

Southern Illinois University
Carbondale, IL

Final Report for:
Summer Research Extension Program

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base
Washington, DC

and

Rome Laboratory

December 1995

TABLE OF CONTENTS

CHAPTERS:

1. Introduction	3
2. Theory for Circular Waveguide to Microstrip Transition	5
3. Theory for Microstrip to Circular Waveguide to Microstrip Transition	8
4. Theory for the Stripline to Circular Waveguide to Stripline Transition	10
5. Microstrip Line Coupled to Aperture Terminated in pmc or pec	13
6. Analysis of the Experimental Results	15
7. Generalization of the Code	28
8. Optimization of CPU Time and Memory	30
9. Stabilized Method for Generating the Maximum Current Along the Microstrip Line	33
10. Conclusion	36

REFERENCES	37
------------	----

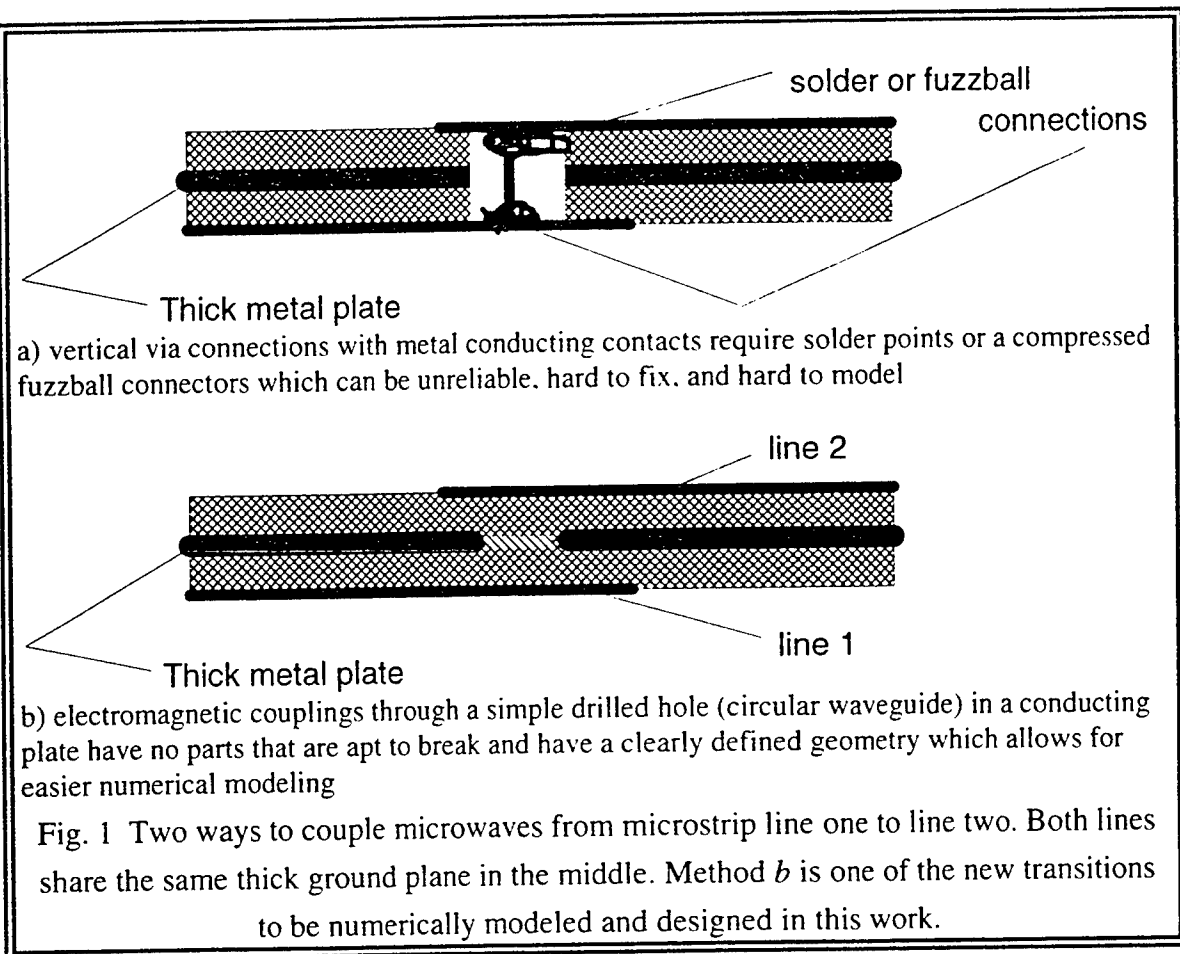
APPENDICES:

A. Basis Function on the Strip with y Dependent Modes	38
B. Method of Moments Equations with y Dependent Modes	39

Chapter 1. Introduction

The project objective was the development of fullwave analyses for solderless coupling between either two microstrip lines or two strip lines through a circular slot in a common, electrically thick ground plane. Similar to the integrated circuit advancement leading to sophisticated computers, there has been a two-decade-long integrated circuit advancement in microwave systems. In state-of-the-art microwave systems, one can find hybrid microwave integrated circuits (MICs) and monolithic microwave integrated circuits (MMICs) which have advantages of being small, lightweight, potentially inexpensive and reliable. Most systems or subsystems have a thick metal ground plane through which the microwave energy must be coupled. The thick metal is used for structural reasons, for dissipating heat from active circuits, or for a sealing boundary between cooled superconducting circuitry and circuitry at ambient temperature. Generally, for coupling rf power vertically through the ground plane, some sort of conducting electrical contact is made, even though soldering or wire bonding can be labor intensive and most expensive [1]. Solder points have other drawbacks as well: they can break and they are not easily modeled at high frequencies where their exact geometry would matter. Some manufacturers have tried fuzzball connectors to make a solderless vertical contact. Fuzzballs are vertical pins with a frayed end that can be pressed against the horizontal metallization to make electrical contact. This design also has the drawbacks of not being reliable and being hard to model, especially at higher frequencies.

One mechanically simple and therefore promisingly reliable alternative to contact vertical coupling is electromagnetic coupling through a hole drilled in the metal plate. The drill-hole is a circular waveguide that may be filled with a material similar to or different from that of the substrates on either side of the metal. See Fig. 1. A similar type of coupling through a small aperture in a thin metallization has been used [2]. Recently an experiment with thick metal plates was conducted [3] with good VSWR results.



The proposed objectives were as follows:

1. Link the subroutines written for the circular waveguide to microstrip transition [4] into a main analysis program using the spectra-domain Green function / Galerkin technique. Test the program for speed and accuracy. Compare the program's value for impedance of a slot-loaded microstrip line to similar results in the literature [2]. (See Chapter 2.)
2. Extend the program to analyze a microstrip to circular waveguide to microstrip transition. This program will be verified against the experiment already conducted [3]. (See Chapter 3.)
3. Do a similar analysis and numerical modeling for stripline to circular waveguide to stripline transition. (See Chapter 4.)
4. Survey the literature for other possible transitions to study and for other cases with which to verify the code. (See Chapter 5.)
5. Report preliminary results of the microstripline and the stripline program at Rome Labs, Hanscom AFB. If necessary, change the prioritization of which generalizations of the code to do next. Possibly do an experiment. (See Chapter 6.)
6. Generalize the code to allow for:
 - a) asymmetric placement of the strip with respect to the coupling hole (See sections 7.1 and 7.2),
 - b) dielectric plug partially filling the waveguide.
 - c) magnetic material partially filling the waveguide,
 - d) biased ferrite partially filling the waveguide, and
 - e) either line can be microstrip or strip line.
7. Optimize the code in terms of CPU time and memory. (See Chapter 8.)

8. Do parameter studies to find designs for best VSWR over an octave or more of frequency.
9. Rewrite the programs from analysis to design tools.
10. Write final report including computer program.

A substantial effort was devoted to meet the proposed objectives. The analytical work for the most of the proposed objectives has been written and the computer algorithms have also been developed.

Chapter 2. Theory for Circular Waveguide to Microstrip Transition

This chapter formulates the analysis of the circular waveguide to microstrip line transition (see Figure 1 in [4]) using the spectral-domain Green-function Galerkin method. The method of moments (MoM) equations were derived during the Summer AFOSR fellowship [4] and the MoM elements were found in the main program. The S-parameters which describe the transition can be calculated after the MoM equations are solved for the expansion coefficients. In this chapter, we consider the waveguide to microstrip transition as a two-port network where port 1 is the microstrip line and port 2 is the circular waveguide. The excitation source at port 2 is the dominant mode of the circular waveguide.

The Method of Moment equations are given as in [4]

$$\overline{\overline{G}}_{0d}^{HJ} \cdot \tilde{\tilde{J}}_s + \overline{\overline{G}}_{00}^{HM} \cdot \tilde{\tilde{M}}_s = \sum_{n=1}^M (V_n^i - V_n^r) \frac{\tilde{\tilde{h}}_n}{Z_n} \quad (2-1)$$

$$\overline{\overline{G}}_{dd}^{EJ} \cdot \tilde{\tilde{J}}_s + \overline{\overline{G}}_{d0}^{EM} \cdot \tilde{\tilde{M}}_s = 0 \quad (2-2)$$

where

$$\begin{aligned} \tilde{\tilde{M}}_s &= -\sum_{n=1}^M (V_n^i + V_n^r) \tilde{\tilde{h}}_n \\ \tilde{\tilde{J}}_s &= \sum_{n=1}^N I_n \tilde{\tilde{J}}_n \end{aligned} \quad (2-3)$$

The corresponding MoM matrix form after Galerkin weighting can be written as

$$\begin{bmatrix} T^{AS} & \eta_o \left(-Y^{AA} + \frac{\delta_{mn}}{Z_n} \right) \\ \frac{Z^{SS}}{\eta_o} & -T^{SA} \end{bmatrix} \begin{bmatrix} \eta I_n \\ V_n^r \end{bmatrix} = \begin{bmatrix} \eta_o \left(Y^{AA} + \frac{\delta_{mn}}{Z_n} \right) \\ T^{SA} \end{bmatrix} \begin{bmatrix} V_n^i \end{bmatrix} \quad (2-4)$$

$$(N+M) \times (N+M) \quad (N+M) \times 1 \quad (N+M) \times (M) \quad (M \times 1)$$

where M is the number of waveguide modes considered and N is the number of rooftop modes considered on the microstrip line [4]. If $M = 3$, then

$$\begin{bmatrix} V_n^i \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \quad (2-5)$$

represents the incident field by an incident dominant mode of the waveguide of unit magnitude and none of the other two waveguide modes (bases) being incident, and

$$\begin{bmatrix} V_n^r \end{bmatrix} = \begin{bmatrix} V_1^r \\ V_2^r \\ V_3^r \end{bmatrix} \quad (2-6)$$

represents the unknown amplitudes of the reflected waveguide modes.

Once all the unknowns are calculated in (2-4), the reflection coefficient seen from port 2 can be expressed as

$$S_{22} = \frac{\text{reflected dominant waveguide mode}}{\text{incident dominant waveguide mode}} \\ = \frac{V_1^r}{V_1^i} \quad \text{at } z = 0 \quad (2-7)$$

where V_1^r and V_1^i are the W.G. dominant mode (1) incident (i) and reflected (r) wave amplitude with phase referenced at $z = 0$. The waveguide to microstrip-line coupling can be expressed as

$$S_{12} = \frac{\text{forward traveling voltage on the strip}}{\text{incident wave from the waveguide (dominant mode)}} \\ = \frac{\sqrt{2}A^+e^{-j\beta x}}{V_1^i} \sqrt{Z_1 Z_2} = \frac{\sqrt{2}A^+}{V_1^i} \sqrt{Z_1 Z_2} \quad \text{at } x = 0 \quad (2-8)$$

where $A^+ = |A^+|e^{j\phi_A}$, Z_2 is the dominant mode wave impedance in the waveguide, and Z_1 is the characteristic impedance of the microstrip line. Refer to the postprocessing section in [4] for finding $|A^+|$, and ϕ_A .

The other S-parameters, S_{11} and S_{21} , which describe the microstrip line to circular waveguide transition can be obtained by making a small modification in (2-1) and (2-2).

The source excitation is moved to port 1 where it is a delta-gap voltage source rather than a waveguide mode. The delta-gap voltage source is located one quarter-wave length from the open end of the microstrip line in order to produce the strongest current along the microstrip line. Then an infinite waveguide(matched) at port 2 is assumed so that no reflections will exist in the waveguide. Similar to (2-1) and (2-2), the MoM equation can be expressed as

$$\overline{\overline{G}}_{0d}^{HJ} \cdot \tilde{\tilde{J}}_s + \overline{\overline{G}}_{00}^{HM} \cdot \tilde{\tilde{M}}_s = \sum_{n=1}^M (-V_n^r) \frac{\tilde{\tilde{h}}_n}{Z_n} \quad (\text{Am}) \quad (2-9)$$

$$\overline{\overline{G}}_{dd}^{EJ} \cdot \tilde{\tilde{J}}_s + \overline{\overline{G}}_{d0}^{EM} \cdot \tilde{\tilde{M}}_s + e^{jk_x x_{\lambda_g} \frac{1}{4}} \left(W \text{sinc} \left(\frac{Wk_y}{2} \right) \right) = 0 \quad (\text{Vm}) \quad (2-10)$$

where W is the width of the microstrip line, and the sinc function term is the delta-gap voltage Fourier transformed, $\tilde{\tilde{M}}_s = -\sum_{n=1}^M V_n^r \tilde{\tilde{h}}_n$, and $\tilde{\tilde{J}}_s$ is as given in (2-3).

By applying the boundary conditions and expanding the unknowns, (2-9) and (2-10) can be written as

$$\overline{\overline{G}}_{0d}^{HJ} \sum_{n=1}^N \eta_o I_n \tilde{\tilde{J}}_n k_o - \eta_o \overline{\overline{G}}_{00}^{HM} \sum_{n=1}^M V_n^r \tilde{\tilde{h}}_n k_o + \eta_o \sum_{n=1}^M V_n^r \frac{\tilde{\tilde{h}}_n}{Z_n} k_o = 0 \quad (\text{V}) \quad (2-11)$$

$$\frac{\overline{\overline{G}}_{dd}^{EJ}}{\eta_o} \cdot \sum_{n=1}^N \eta_o I_n \tilde{\tilde{J}}_n k_o - \overline{\overline{G}}_{d0}^{EM} \cdot \sum_{n=1}^M V_n^r \tilde{\tilde{h}}_n k_o = -e^{jk_x x_{\lambda_g} \frac{1}{4}} \left(W \text{sinc} \left(\frac{Wk_y}{2} \right) \right) \quad (\text{V}) \quad (2-12)$$

After Galerkin weighting, equation (2-11) and (2-12) can be written in matrix form as

$$\begin{bmatrix} T^{AS} & \eta_o \left(-Y^{AA} + \frac{\delta_{mn}}{Z_n} \right) \\ \frac{Z^{SS}}{\eta_o} & -T^{SA} \end{bmatrix} \begin{bmatrix} \eta_o I_n \\ V_n^r \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ -k_o \\ 0 \end{bmatrix} \quad (\text{V}) \quad (2-13)$$

$$(N+M) \times (N+M) \quad (N+M) \times 1 \quad (N+M) \times 1$$

where T^{AS} , Z^{SS} , T^{SA} , and Y^{AA} are given in [4], and the non-zero term on the right side corresponds to the location of the delta-gap source on the microstrip line. There are N

rooftop modes along the strip with equal spacing from each other. Upon solving the unknowns, the reflection coefficient seen from port 1 as well as the coupling from the microstrip line to the waveguide can be found.

$$S_{11} = \frac{A^+ \operatorname{Re}^{j\phi} e^{j\beta x}}{A^+ e^{-j\beta x}} = \operatorname{Re}^{j(\phi+2\beta x)} = \operatorname{Re}^{j\phi} \quad \text{at } x = 0 \quad (2-14)$$

And by using (F-1c) and (E-17)

$$\begin{aligned} S_{21} &= \frac{\text{dominant mode into the waveguide}}{\text{backward traveling voltage on the strip}} \\ &= \frac{V_1^r}{\sqrt{2} A^+ e^{-j\beta x} \operatorname{Re}^{j\phi}} \frac{1}{\sqrt{Z_1 Z_2}} = \frac{V_1^r}{\sqrt{2} A^+ \operatorname{Re}^{j\phi}} \frac{1}{\sqrt{Z_1 Z_2}} \quad \text{at } x = 0 \end{aligned} \quad (2-15)$$

where R and ϕ are the magnitude and phase of the current reflection coefficient respectively, and V_1^r is the dominant waveguide mode traveling in the waveguide in $-\hat{z}$ direction. Again refer to postprocessing section in [4] for finding $|A^+|$, ϕ_A , R , and ϕ .

Chapter 3. Theory for Microstrip to Circular Waveguide to Microstrip Transition

This chapter extends the analysis done in the previous chapter which described fully the two-port transition of circular waveguide to microstrip line. Here the analysis of a microstrip to circular waveguide to microstrip transition is considered.

The geometry analyzed here is:

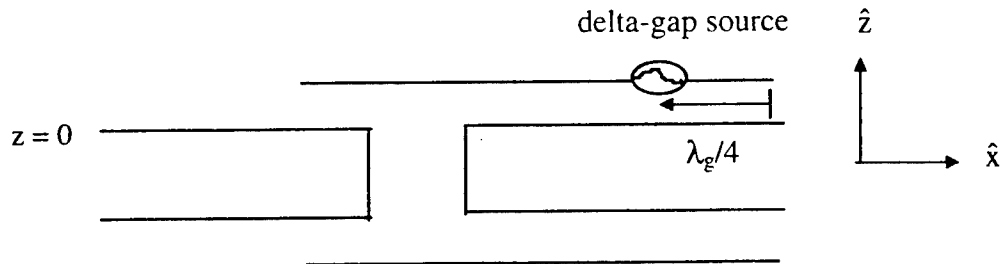


Fig. 3-1 Side view of microstrip to microstrip transition in symmetry

The transition can be described by cascading three two-port networks two of which were described by the S parameters in the previous chapter. Refer to the Fig. 3-2 for three cascaded two-port networks. The center two-port network is a section of waveguide transmission line.

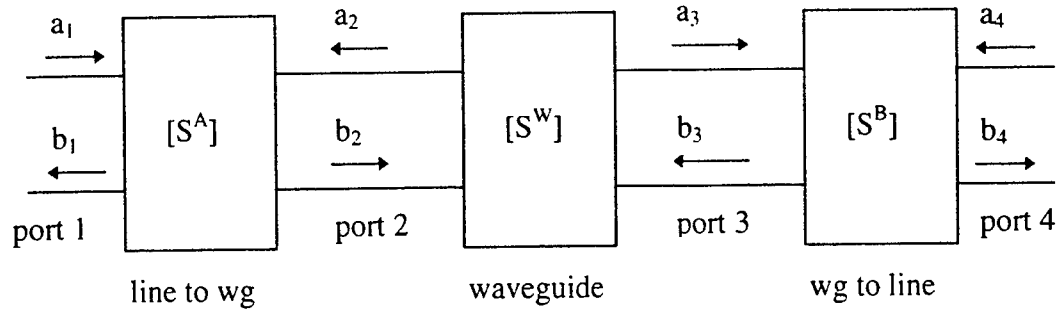


Fig. 3-2 Cascaded three two-port networks

These three cascaded two-port networks result in one two-port network as shown below

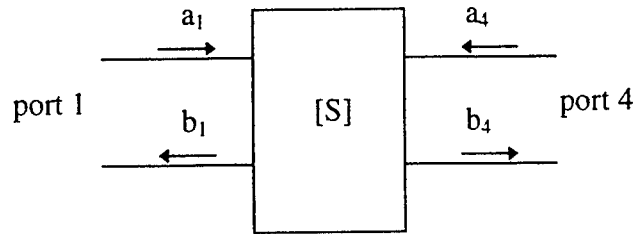


Fig. 3-3 Two-port network

In general

$$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} S_{11}^A & S_{12}^A \\ S_{21}^A & S_{22}^A \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} \quad \text{for line to wg} \quad (3-1)$$

$$\begin{bmatrix} b_4 \\ b_3 \end{bmatrix} = \begin{bmatrix} S_{44}^B & S_{43}^B \\ S_{34}^B & S_{33}^B \end{bmatrix} \begin{bmatrix} a_4 \\ a_3 \end{bmatrix} \quad \text{for wg to line} \quad (3-2)$$

$$\begin{bmatrix} a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} S_{22}^W & S_{23}^W \\ S_{32}^W & S_{33}^W \end{bmatrix} \begin{bmatrix} b_2 \\ b_3 \end{bmatrix} \quad \text{for wg to wg} \quad (3-3)$$

The S-parameter elements in (3-3) are,

$$S_{22}^W = S_{33}^W = 0 \quad (3-4)$$

$$S_{23}^W = e^{j\beta_n l} \text{ and } S_{32}^W = e^{-j\beta_n l}, \quad (3-5)$$

where l is the exact length of the circular waveguide. With some algebraic computation the cascaded S-parameters can be written as

$$\begin{bmatrix} b_1 \\ b_4 \end{bmatrix} = \begin{bmatrix} S_{11} & S_{14} \\ S_{41} & S_{44} \end{bmatrix} \begin{bmatrix} a_1 \\ a_4 \end{bmatrix} \quad (3-6)$$

where these four S parameters are defined in terms of the S parameters of the individual components composing this network as shown below.

$$S_{11} = \frac{S_{11}^A + [S_{12}^A]^2 [S_{32}^W]^2 S_{33}^B - [S_{32}^W]^2 S_{33}^B S_{22}^A S_{11}^A}{1 - [S_{32}^W]^2 S_{33}^B S_{22}^A} \quad (3-7)$$

$$S_{14} = \frac{S_{12}^A S_{32}^W S_{34}^B}{1 - [S_{32}^W]^2 S_{33}^B S_{22}^A} \quad (3-8)$$

$$S_{41} = \frac{S_{43}^B S_{32}^W S_{21}^A}{1 - [S_{32}^W]^2 S_{33}^B S_{22}^A} \quad (3-9)$$

$$S_{44} = \frac{S_{44}^B + S_{43}^B S_{32}^W S_{22}^A S_{23}^W S_{34}^B - S_{32}^W S_{22}^A S_{23}^W S_{33}^B S_{44}^B}{1 - [S_{32}^W]^2 S_{33}^B S_{22}^A} \quad (3-10)$$

S^A and S^B are defined as microstrip line to waveguide transition and waveguide to microstrip line transition as described in (2-7), (2-8), (2-14), and (2-15).

Rotation of one line with respect to the other are accounted for by simply rotating the waveguide. (See section 7.2.)

Chapter 4. Theory for the Stripline to Circular Waveguide to Stripline Transition

The effectiveness of a method-of-moments (MoM) solution depends on a judicious choice of basis function. These basis functions should incorporate as closely as possible the physical conditions of the actual axial and transverse currents on the stripline. The transverse current is usually small compared to the longitudinal current and its magnitude is proportional to the strip width. Therefore, as the strip width becomes wider it is necessary to put transverse current modes with edge conditions on the stripline to improve accuracy [5], [7], [14].

4.1 Basis Functions and MoM Equations on the Strip with y dependent modes

First, instead of using longitudinal modes with constant y (transverse) variation as in Appendix C in [4], one can use longitudinal modes with a y-dependent edge condition on the stripline. First, we considered a simple geometry where the stripline's left edge is centered at $(x_s, 0)$. The line has width W and length L. The electric surface current on the strip is expanded in triangular-pulse functions with a few y-dependent modes which have edge conditions given in Appendix A.

By using these current modes, the Y^{AA} submatrix in the MoM equation in Appendix B stays the same as shown in Appendix D in [4], but submatrices T^{SA} , T^{AS} , and Z^{SS} should be changed as shown in Appendix H. These current modes can be also applied for the microstrip problems in chapter 2, 3, and 5. After solving the MoM equations (2-4) in chapter 2 with using submatrices T^{SA} , T^{AS} , and Z^{SS} in Appendix B, S_{22} and S_{12} , which describe the circular waveguide to the microstrip line transition can be obtained by (2-7) and (2-8). S_{11} and S_{21} , which describe the microstrip line to the circular waveguide transition can be obtained by (2-13), (2-14), and (2-15). S_{11} for the microstrip to microstrip through circular waveguide transition can be obtained by (5-2), (5-6), (5-7), and (5-8). The results for microstrip problem with y-dependent modes list in chapter 6.

4.2 Transverse and Longitudinal Directed Current Basis Functions with Edge Conditions on the Stripline and the MoM Equations

Now consider both longitudinal and transverse current modes. For our structure, the electric surface current on the strip is expanded in triangular pulse functions with edge conditions for longitudinal and transverse current components as in [5], [7]. A Chebyshev polynomial of the first kind (T_{2i_p}) is used to expand the longitudinal current modes and a Chebyshev polynomial of the second kind (U_{2k_p-1}) is used to expand the transverse current modes. The square-root factors is also used for the anticipated edge behavior. The line has width W and length L . Again, the striplines's left edge is centered at $(x_s, 0)$. By using these current modes, the MoM equations and the submatrix Y^{AA} remain the same as shown Appendix D in [4], but the submatrices T^{SA} , T^{AS} , and Z^{SS} are changed. A closed-form expression for transverse current distribution is also found in [14]. However, we don't derive the submatrices of the MoM equation for this case because our narrow stripwidth make zero current distribution for a closed-form expression in [14].

4.3 Green's Function for Stripline Problem

The Green function required for the stripline problem is different from that used in the microstripline problem. The stripline has a ground plane above and below the material containing the strip. A derivation of the necessary Green functions has been done.

Chapter 5. Microstrip Line Coupled to Aperture Terminated in pmc or pec

For a simple symmetric geometry, one way of calculating S_{11} for the microstrip to microstrip through circular waveguide transition is by using even and odd analysis. The simple geometry has odd symmetry about the center of the length of waveguide. Even analysis is done by open circuiting the center of the waveguide. Odd analysis is done by short circuiting the center of the waveguide. The total S_{11} is then the superposition of the two results. This result can be used to verify the S_{11} obtained in chapter 3.

At the aperture

Refer to the Fig. 5-1 for coordinate system for the waveguide reflection coefficient.

The reflection coefficients of waveguide modes at an open or short load (at $z = -l$) can be written as

$$\Gamma_n(-l) = \pm 1 \quad (+1 \text{ for open circuit and } -1 \text{ for short circuit}) \quad (5-1)$$

Incident and reflected wave in terms of (5-1) at $z = 0$ can be written as

$$\Gamma_n(0) = \frac{V_n^i}{V_n^r} = \begin{cases} -e^{-2j\beta_n l} & \text{for short circuit } (Z_L = 0) \\ +e^{-2j\beta_n l} & \text{for open circuit } (Z_L = \infty) \end{cases} \quad (5-2)$$

where β_n is defined in (E-8).

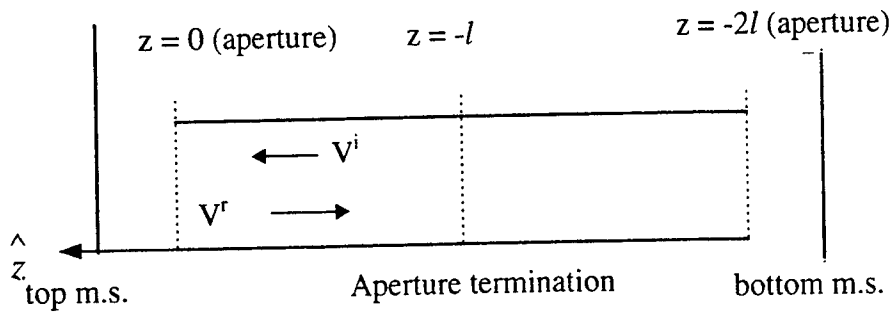


Fig. 5-1 Aperture terminated in pmc or pec

Accounting for this waveguide termination as in (5-2), the magnetic field integral equation (MFIE) (2-1) becomes

$$\overset{=}{G}_{0d}^{HJ} \bullet \sum_{n=1}^N I_n \tilde{\tilde{J}}_n + \overset{=}{G}_{00}^{HM} \bullet - \sum_{n=1}^M (\Gamma_n(0) + 1) V_n' \tilde{\tilde{h}}_n = \sum_{n=1}^M (\Gamma_n(0) - 1) V_n' \frac{\tilde{\tilde{h}}_n}{Z_n} \quad (5-3)$$

On the microstrip

The electric field integral equation (EFIE) comes from (2-2) and (2-3) but modified to include the delta-gap voltage source on the microstrip. It is

$$\overset{=}{G}_{dd}^{EJ} \bullet \sum_{n=1}^N I_n \tilde{\tilde{J}}_n + e^{\frac{jk_z x_{\lambda_g}}{4}} \left(W \operatorname{sinc} \left(\frac{Wk_y}{2} \right) \right) + \overset{=}{G}_{d0}^{EM} \bullet - \sum_{n=1}^M (V_n^i + V_n') \tilde{\tilde{h}}_n = 0. \quad (5-4)$$

Substitution of (5-2) into (5-4) yields

$$\overset{=}{G}_{dd}^{EJ} \bullet \sum_{n=1}^N I_n \tilde{\tilde{J}}_n + e^{\frac{jk_z x_{\lambda_g}}{4}} \left(W \operatorname{sinc} \left(\frac{Wk_y}{2} \right) \right) + \overset{=}{G}_{d0}^{EM} \bullet - \sum_{n=1}^M (\Gamma_n(0) + 1) V_n' \tilde{\tilde{h}}_n = 0 \quad (5-5)$$

After the Galerkin weighting, the MoM in matrix form for (5-3) and (5-5) can be written as

$$\begin{bmatrix} T^{AS} & \eta_o \left(-Y^{AA} (\Gamma_n(0) + 1) + \frac{\delta_{nn}}{Z_n} (\Gamma_n(0) - 1) \right) \\ \frac{Z^{SS}}{\eta_o} & -T^{SA} (\Gamma_n(0) + 1) \end{bmatrix} \begin{bmatrix} \eta_o I_n \\ V_n' \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ -k_o \\ 0 \end{bmatrix} \quad (5-6)$$

(N+M)x(N+M) (N+M)x1 (N+M)x1

The excitation is located one quarter wavelength from the edge of the strip.

The even and odd S_{11} can be found by using (5-6) with (5-2). By superposing the two results, the total S_{11} can be written as

$$S_{11} = \frac{S_{11}^{o.c.} + S_{11}^{s.c.}}{2} \quad (5-7)$$

$$= \frac{R^{o.c.} e^{j\phi^{o.c.}} + R^{s.c.} e^{j\phi^{s.c.}}}{2}. \quad (5-8)$$

Chapter 6. Analysis of the Experimental Results

The computer simulations for the various waveguide terminations were investigated. A series of reflection coefficient versus frequency plots for the circular waveguide terminated with matched load, and perfect magnetic conductor (pmc) and perfect electric conductor (pec) are presented in Fig. 6-1 to Fig. 6-3. The thickness of the ground plate used for pmc and pec termination was 12.19mm while the match load was simulated by assuming an infinitely long waveguide, thus having no reflection from the termination. The frequency range used in all the figures is 4-6 GHz. Fig. 6-1 presents the reflection coefficient versus frequency for the matched load using several different techniques to calculate it. The top two graphs are the magnitude of reflection coefficient results obtained by using the postprocessing routine with and without interpolation which samples the unknown current coefficients along the strip and extracts the forward and backward traveling waves. The magnitude of reflection coefficients in the bottom two graphs were obtained by finding the maximum and minimum values of the current standing wave generated on the microstrip. Each graph is compared to the data provided by Marat who had already conducted the same simulation and was able to match fairly well against the actual experimental case. Most of the results were in good agreement except the case where the postprocessing routine without interpolation was used. The large error in this case resulted probably due to the fact that the postprocessing routine requires an even number of sample points within the half wavelength period, and not all frequencies were able to meet such requirement (some frequencies had odd number of sample points within the half wavelength period). In such case, total number of sample points were adjusted by either adding or subtracting by one sample point. The adjustment in the sampling points may have led to slightly inaccurate current calculation, thus an error in reflection coefficient calculation. To reduce such an error we need more sampling points. Thus, we interpolate unknown current coefficients along the strip with a cubic spline. Fig. 6-3 presents the reflection coefficient calculated for microstrip to waveguide to microstrip transition using even and odd analysis. The circular waveguide was terminated with both pmc and pec, and reflection coefficients were calculated for each termination. The total reflection coefficient was then calculated by superposing the two results. In Fig. 6-1, the reflection coefficient was calculated by using longitudinal modes with constant y (transverse) variation. Fig. 6-2 also shows the reflection coefficient versus frequency as shown Fig. 6-1 but with y -dependent modes.

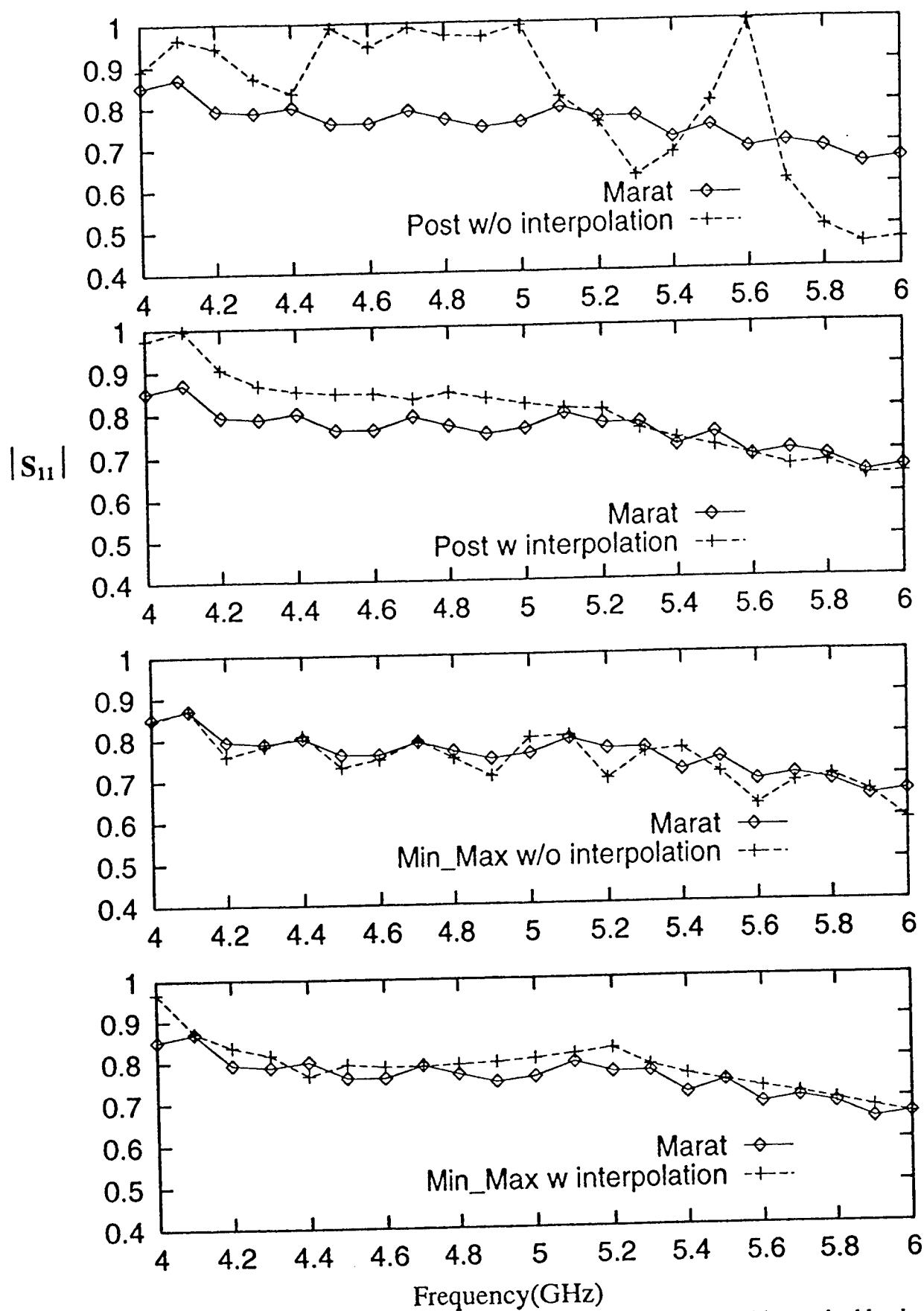


Fig. 6-1 The reflection coefficient for the circular WG terminated with matched load.

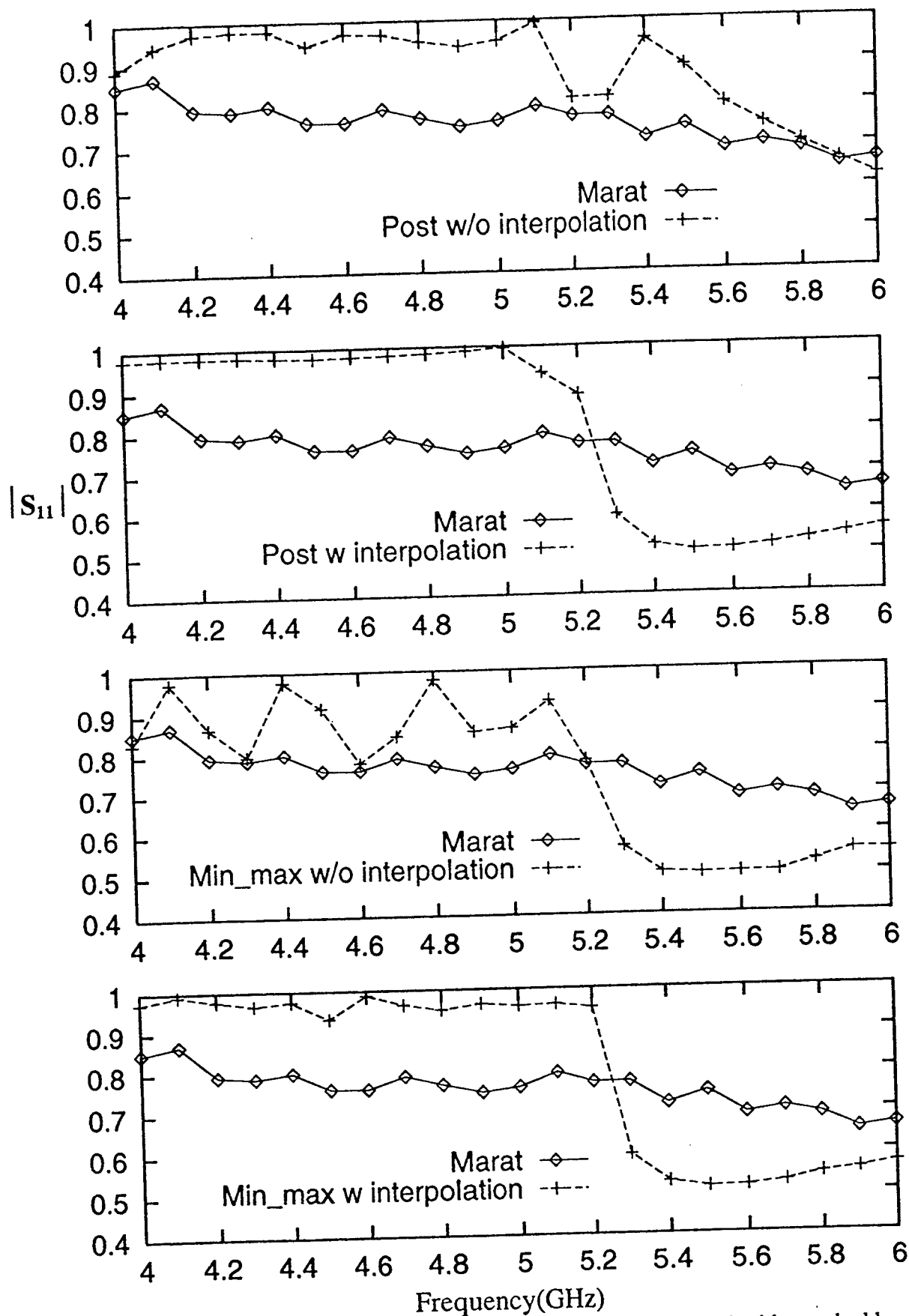


Fig. 6-2 The reflection coefficient for the circular WG terminated with matched load by using y-dependent modes

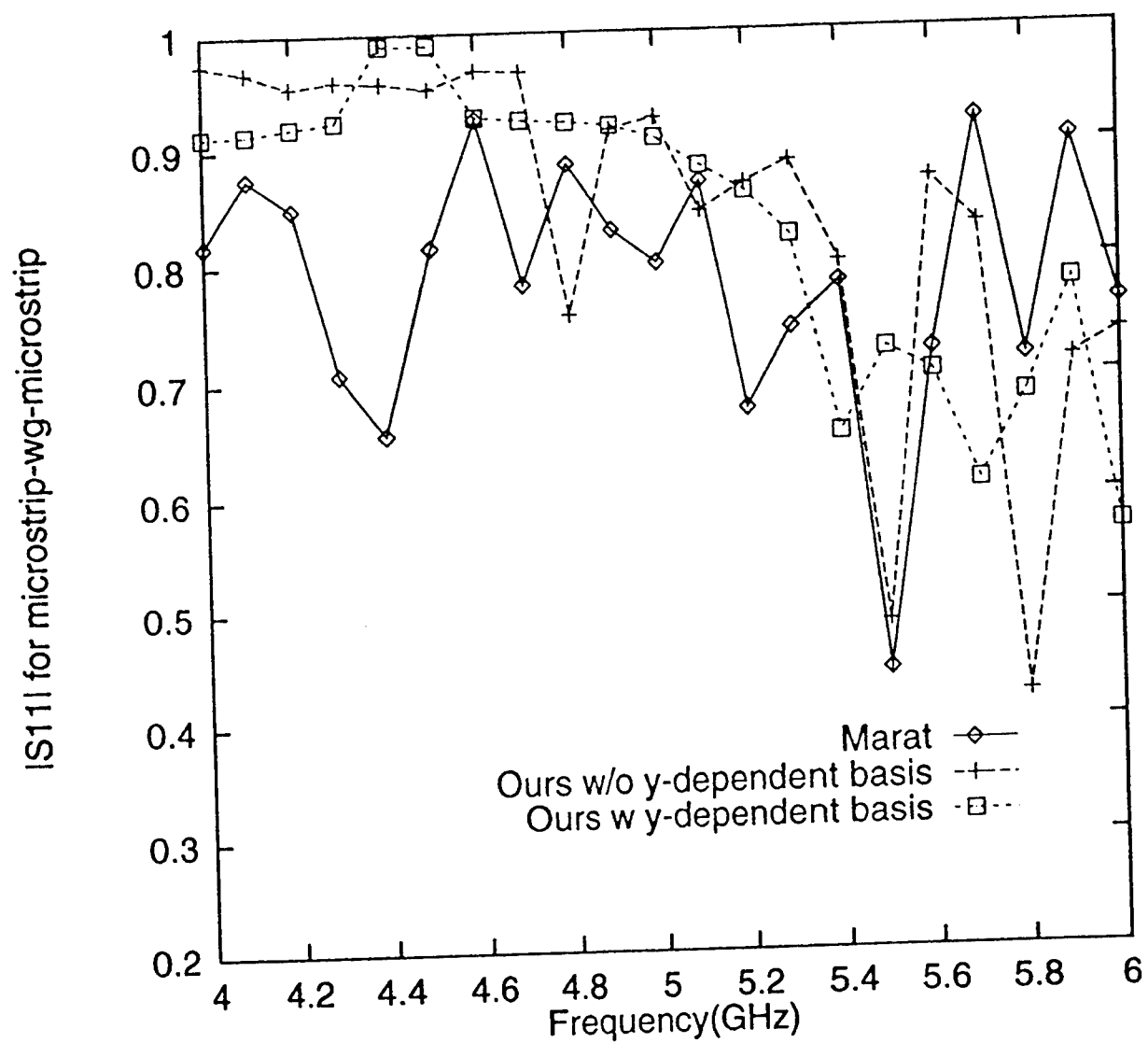


Fig. 6-3 The reflection coefficient for microstrip to WG to microstrip

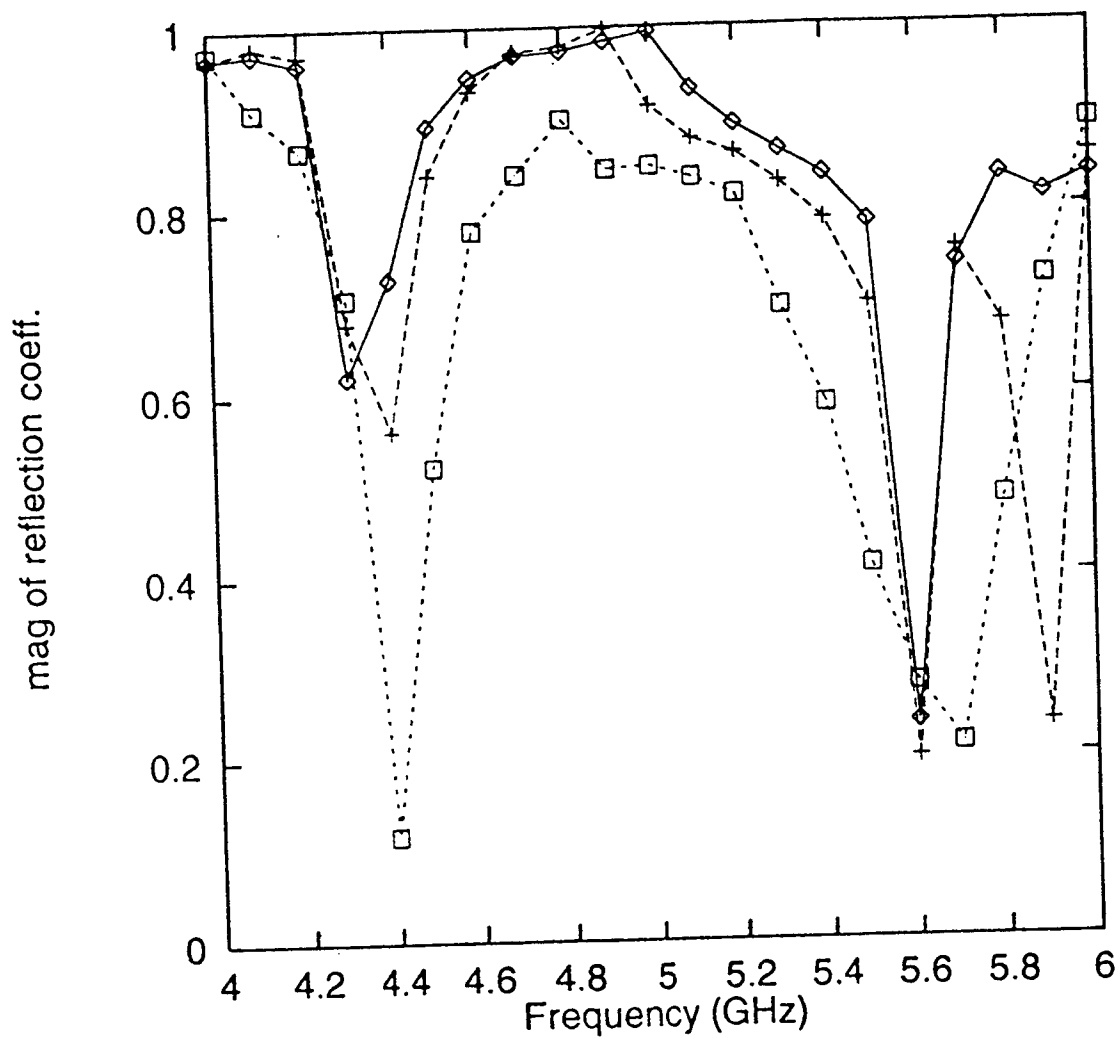


Fig. 6-4 The reflection coefficient for MS-WG-MS with different stub length. \square --- $x_s = -9.51\text{mm}$, $+$ --- $x_s = -13.51\text{mm}$, \diamond --- $x_s = -16.51\text{mm}$

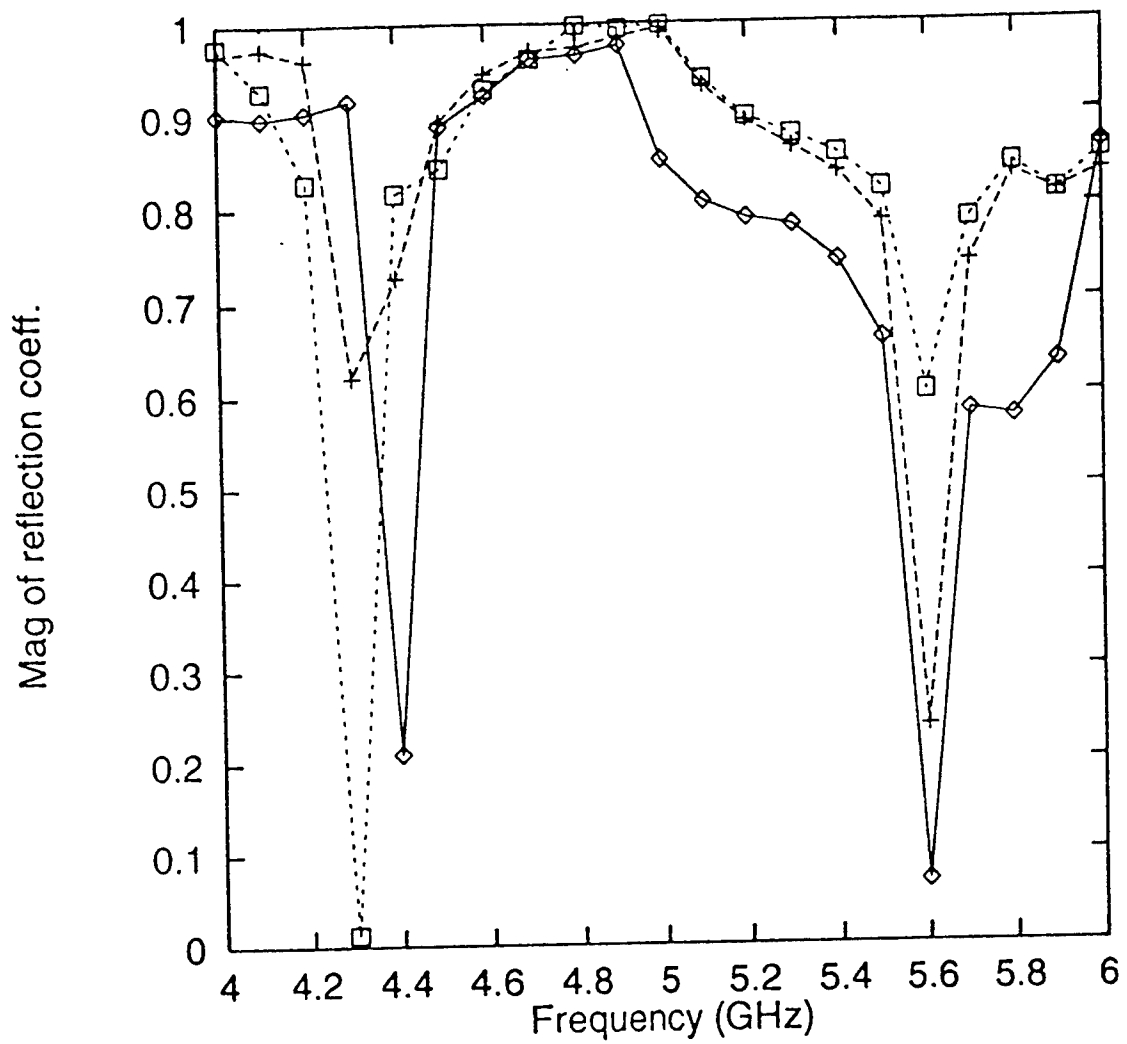


Fig. 6-5 The reflection coefficient for MS-WG-MS with different substrate thickness. \square --- $d = 0.57\text{mm}$, $+$ --- $d = 1.57\text{mm}$, \diamond --- $d = 2.57\text{mm}$

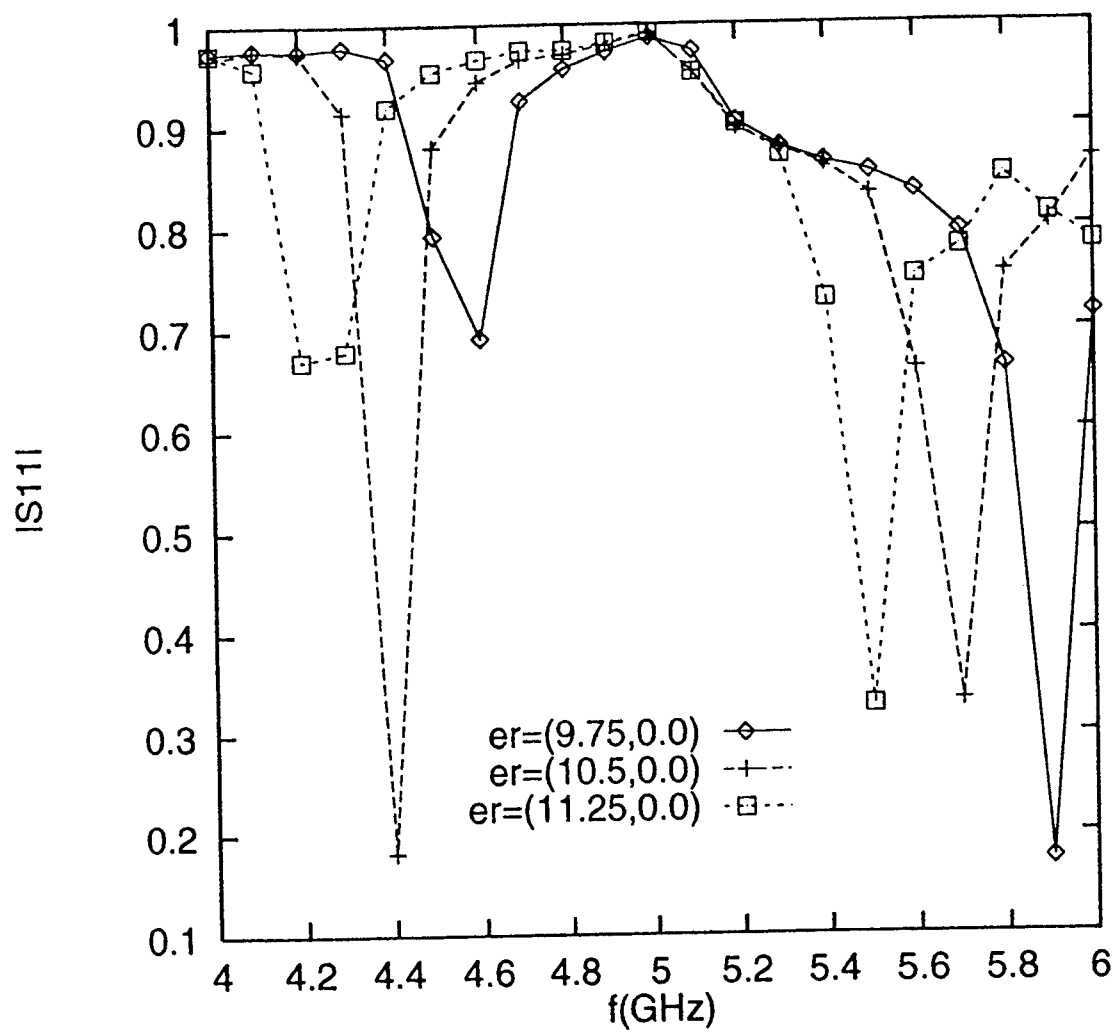


Fig. 6-6 The reflection coefficient for microstrip to WG to microstrip with WG filled with various dielectric constants and by using longitudinal modes with constant y variation.

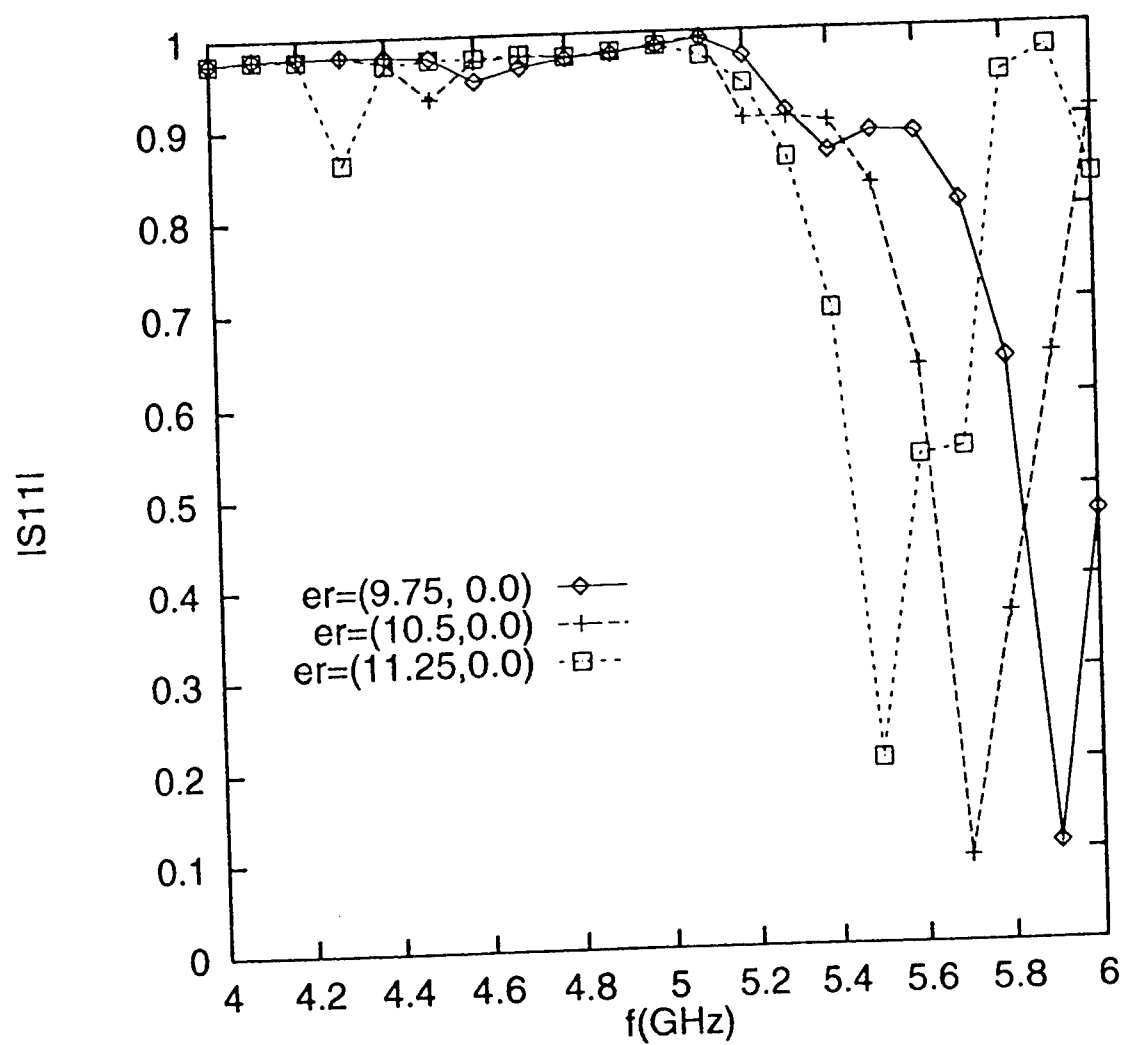


Fig. 6-7 The reflection coefficient for microstrip to WG to microstrip with WG filled with various dielectric constants and by using longitudinal y-dependent modes

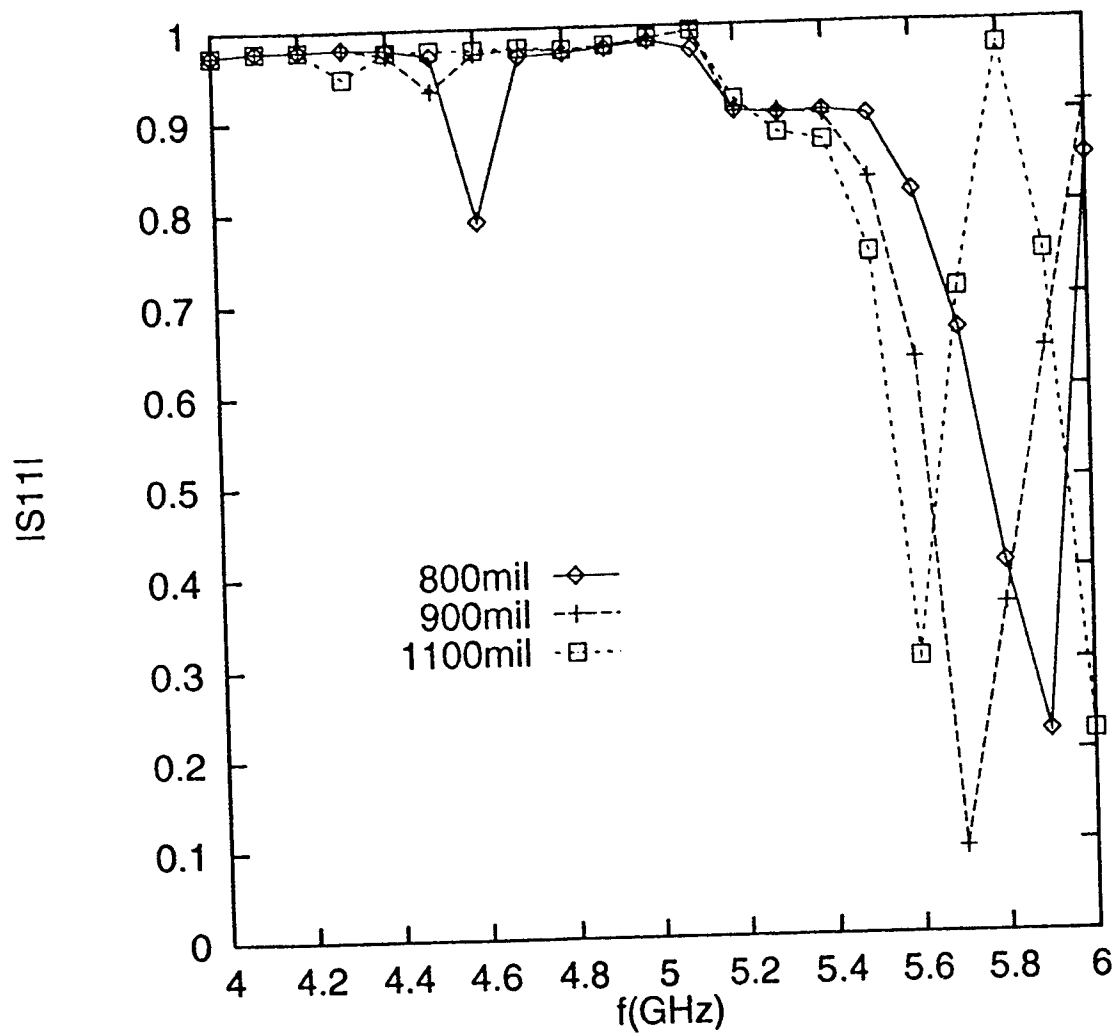


Fig. 6-9 The reflection coefficient for microstrip to WG to microstrip with various WG lengths and by using longitudinal y-dependent modes

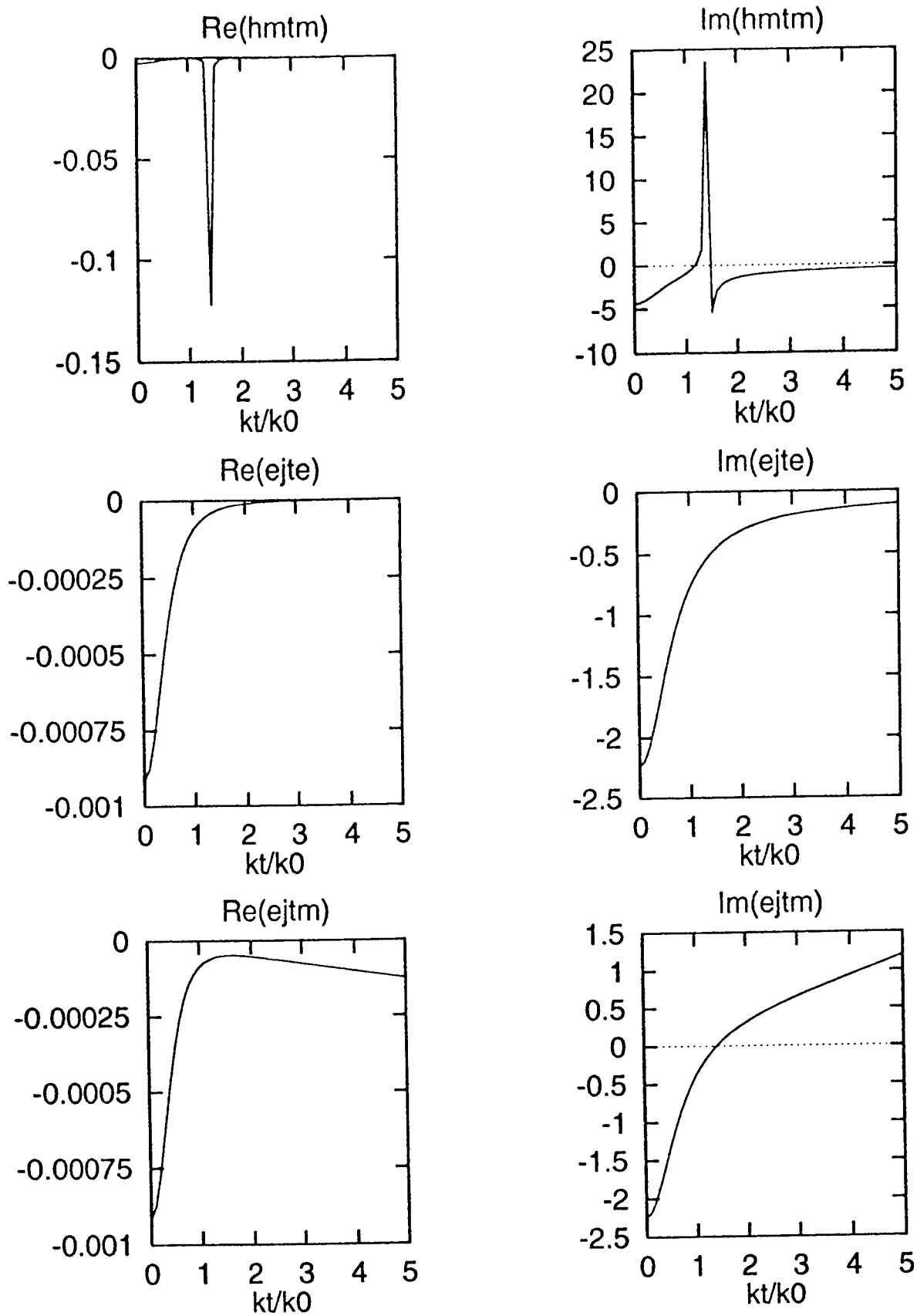


Fig. 6-11 Real and imaginary parts of the normalized Green function for stripline problem: $hmtm$, $ejte$, $ejtm$

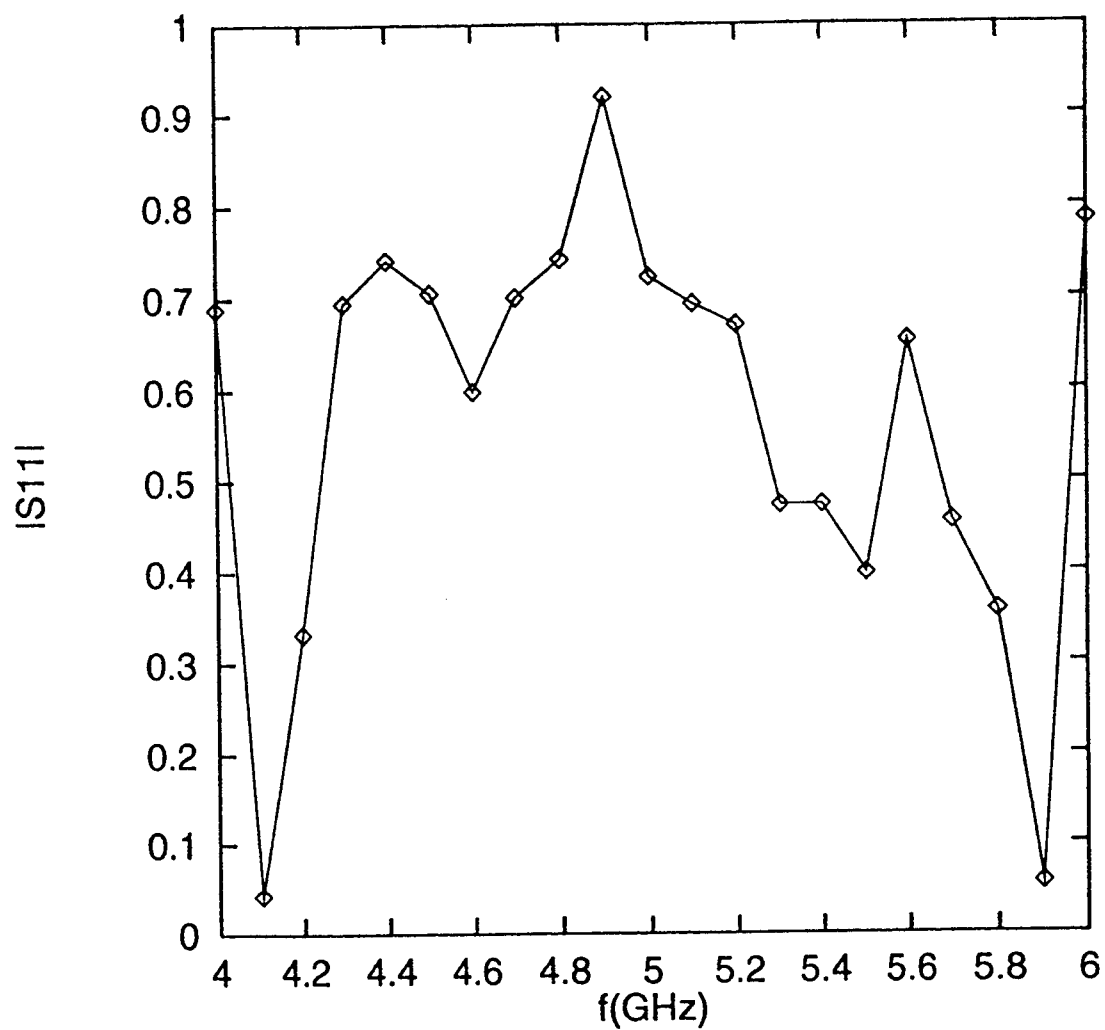


Fig. 6-12 The reflection coefficient for stripline to WG to stripline.
The thickness above and below stripline = 62 mil

Chapter 7. Generalization of the Code

7.1 Asymmetric Placement of the Strip with respect to the Coupling Hole

In [4], the general circular waveguide to microstripline transition is studied with $y_s=0$. The stripline to circular waveguide to stripline transition is also analyzed with $y_s=0$ in chapter 4. Here we consider that the microstrip or strip line's left edge is not centered at $(x_s, 0)$, and rederive the submatrices of the MoM matrix equation. The matrices Y^{AA} and Z^{SS} stay the same as in [4] but the matrix T^{SA} is changed. The matrix T^{SA} for the circular waveguide to microstripline transition with $y_s \neq 0$ has been calculated. The matrix T^{SA} for the stripline to circular waveguide to stripline transition where the strip line's left edge is centered at (x_s, y_s) with y dependent longitudinal modes without transverse mode on the strip has been calculated. The matrix T^{SA} for transverse and longitudinal current components on the stripline has been calculated.

7.2 Asymmetric Placement of the Microstrip with Respect to y-axis

A microstrip to waveguide to microstrip transition with the two microstrip lines oriented 180 degrees apart is considered in this chapter. Refer to the Fig. 7-1 for the geometry being analyzed.

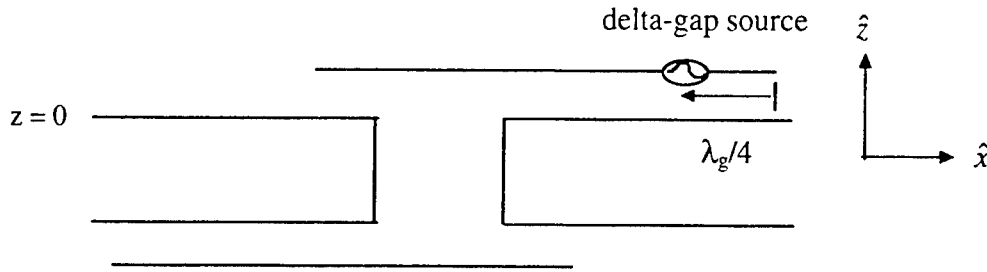


Fig. 7-1 Rotation of the bottom microstrip w.r.t. the top microstrip

A simple way of achieving such configuration is by rotating the waveguide by 180 degrees (see Fig. 7-2). The S-parameters of three two port devices can be derived similarly as in chapter 3. The S-parameters in (3-1) and (3-2) remain the same while S_{23}^w and S_{32}^w in (3-3) are replaced with minus sign in front [3]. This is because that 180

degrees rotation of the waveguide is equivalent to the variation in the phase of a propagating wave by 180 degrees.

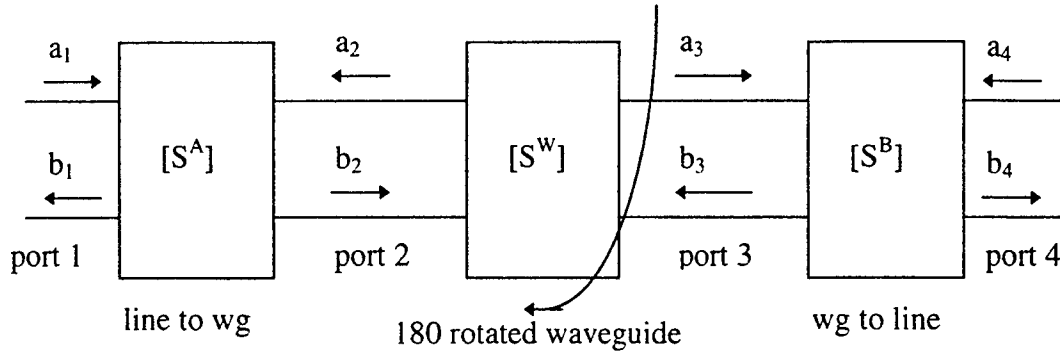


Fig. 7-2 Cascaded three two-port networks with the waveguide rotated by 180 degrees

Upon the waveguide rotation, the elements of the S-parameter matrix in (3-3) are modified to

$$\begin{bmatrix} a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} S_{22}^W & -S_{23}^W \\ -S_{32}^W & S_{33}^W \end{bmatrix} \begin{bmatrix} b_2 \\ b_3 \end{bmatrix} \quad \text{for wg to wg} \quad (7-1)$$

Using (3-4), (3-5), and (7-1), the S-parameters of cascaded three two-port networks with the waveguide rotated by 180 degrees can be written as

$$\begin{bmatrix} b_1 \\ b_4 \end{bmatrix} = \begin{bmatrix} S_{11} & S_{14} \\ S_{41} & S_{44} \end{bmatrix} \begin{bmatrix} a_1 \\ a_4 \end{bmatrix} \quad (7-2)$$

where

$$S_{11} = \frac{S_{11}^A + [S_{12}^A]^2 [S_{32}^W]^2 S_{33}^B - [S_{32}^W]^2 S_{33}^B S_{22}^A S_{11}^A}{1 - [S_{32}^W]^2 S_{33}^B S_{22}^A} \quad (7-3)$$

$$S_{14} = \frac{-S_{12}^A S_{32}^W S_{34}^B}{1 - [S_{32}^W]^2 S_{33}^B S_{22}^A} \quad (7-4)$$

$$S_{41} = \frac{-S_{43}^B S_{32}^W S_{21}^A}{1 - [S_{32}^W]^2 S_{33}^B S_{22}^A} \quad (7-5)$$

$$S_{44} = \frac{S_{44}^B + S_{43}^B S_{32}^W S_{22}^A S_{23}^W S_{34}^B - S_{32}^W S_{22}^A S_{23}^W S_{33}^B S_{44}^B}{1 - [S_{32}^W]^2 S_{33}^B S_{22}^A} \quad (7-6)$$

Chapter 8. Optimization of CPU Time and Memory

The summary of steps taken to make the program which calculates the MoM matrix elements run faster is shown in the table below.

COMPILING OPTIONS USED	PROGRAM CHANGES MADE	EXECUTION TIME
1. f77 -pg	M=1, N=2, kt integrated from 0 to 280*k0	2.8 hours
2. “	ca=cos(alpha), sa=sin(alpha)	2.2 hours
3. f77 -pg -fast	none	0.98 hour
4. f77	none	1.2 hours
5. f77 -fast	extra routine and write statement to look at n distribution	0.99 hour
6. f77 -fast	started n at 300, requested absolute error of 1e-4 rather than 1e-8	0.36 hour
7. f77 -fast	write statement deln, # times deln was chosen	0.37 hour
8. f77 -fast	n starting at 1490 rather than at 300	0.12 hour
9. f77	greens subroutine is split into gej, ghj, and ghm. alphai is split into alphai and dalphai for the sake of double cases in tintgd. Subroutine tdoub with its 5 cases replaces ti1 with 3 cases and ti2 with 5 cases. htrans is split into alanal(alpha,phi,M) and htrans(kt). A lot of parameters such as deln,n,zintgd are written to the screen.	0.13 hour
10. f77 -fast	“	0.11 hour
11. f77 -fast	write to screen less	0.10 hour
12. f77 -fast speed.f -lcmf -o speed	compiled with library that was compiled with fast option	0.10 hour
13.”	Subroutine htrans was modified so that kta is now an array and is calculated in its own loop and is not in the loop over i=1,mnum. Also, the unused quantities of kr, etar, krak0, and odd are not defined or calculated	0.10 hour
14.”	Subroutines zss, zintgd, zi1c, and zi2s were changed so that 0, p-q is passed rather than p,q. This eliminates one subtraction from zi1c and one from zi2s which are both called quite often (thousands of times)	0.10 hour
15. “	Call zi1c and zi2s together in one subroutine zdoub rather than in two functions.	0.06 hour
16. “	M=2, N=2.	0.26 hour

	<pre> tint13 = tint13*sin(alpha*malphap))*sa Replaced in zdoub the following: c Start assembling zi1c and zi2s. Here is the old method: c zi1c = ca*ca*cos(xkx*h0*delpq) zi2s = sa*sa*cos(xkx*h0*delpq) sincx = sin(argx)/argx sincx4 = sincx**4 sincy = sin(argy)/argy sincy2 = sincy*sincy zi1c = zi1c*sincx4*sincy2 zi2s = zi2s*sincx4*sincy2 with: c Start assembling zi1c and zi2s. Here is the new method: c zi1c = cos(xkx*h0*delpq) sincx = sin(argx)/argx sincx4 = sincx**4 sincy = sin(argy)/argy sincy2 = sincy*sincy zi1c = zi1c*sincx4*sincy2 zi2s = zi1c*sa*sa zi1c = zi1c*ca*ca </pre>	
19. f77 -pg speed.f -lcmf -o speedpg	“	0.42 hour
20. f77 -fast speed.f -lcmf -o speed	Corrected alanal and htrans subroutines.	0.19 hour

Then the entire program was compiled with f77 ver0401.f -lcmf and run with N=40 and M=2. The user time was 7:30:30. Next, a good approximation was implemented: any elements of Z^{SS} or of T^{AS} whose indices were separated by 25 or more were set to zero. This ignores the interaction for any points greater than one half wavelength away. The user time was thus reduced to 2:28:15 or only 40% of the previous fastest time or about a 99.8% reduction from the original time.

Chapter 9. Stabilized Method for Generating the Maximum Current Along the Microstrip Line

To produce a maximum current magnitude along the microstrip line, the delta-gap voltage source is placed at exactly one quarter wavelength from the open end of the line. The point exactly one quarter wavelength from the end will not in general be coincidentally at the center of one of the rooftop basis functions. Usually it is at a point where two roof top functions are nonzero. When a single rooftop basis function is used to locate the source excitation, the nearest roof-top-mode center was chosen as the excitation point. Moving the excitation off from exactly one-quarter wavelength will cause a reduction in current magnitude on the line. Thus, for maximum current, two roof tops are used to give the source excitation at exactly quarter wavelength from the end of the line. This can be accomplished by calculating the relative amplitudes of the two roof tops that overlap the source point [3]. Thus, the right side of the equation in (2-13) and (5-6) will have two non-zero elements where the non-zero elements can be found from [4]. Rewriting (2-13) and (5-6) with two non-zero elements become

$$\begin{bmatrix} T^{AS} & \eta_o \left(-Y^{AA} + \frac{\delta_{mn}}{Z_n} \right) \\ \frac{Z^{SS}}{\eta_o} & -T^{SA} \end{bmatrix} \begin{bmatrix} \eta_o I_n \\ V_n^r \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ -A \\ -B \\ 0 \end{bmatrix} \quad (9-1)$$

and

$$\begin{bmatrix} T^{AS} & \eta_o \left(-Y^{AA} (\Gamma_n(0) + 1) + \frac{\delta_{mn}}{Z_n} (\Gamma_n(0) - 1) \right) \\ \frac{Z^{SS}}{\eta_o} & -T^{SA} (\Gamma_n(0) + 1) \end{bmatrix} \begin{bmatrix} \eta_o I_n \\ V_n^r \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ -A \\ -B \\ 0 \end{bmatrix} \quad (9-2)$$

where A and B are the relative amplitude of the two roof tops with maximum being one. Relative amplitude of the two roof tops can be seen in Fig. 9.

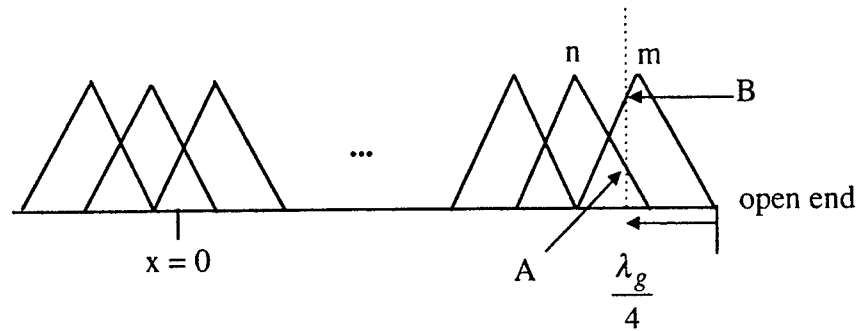


Fig. 9 Delta gap voltage source location

The n and m in the Fig. 9 are the two roof tops that overlap at the source point. The relative amplitude of the two roof tops can be calculated by knowing the guided wavelength at the operating frequency, the strip length and the number of roof tops being used. By using the triangular-pulse functions given in [4], A and B can be calculated as

$$A = 1 - \frac{|x - x_s - nh|}{h} \quad (9-3)$$

$$B = 1 - \frac{|x - x_s - mh|}{h} \quad (9-4)$$

where h is the segment between two adjacent roof tops, x_s is stub length, and x is the quarter wavelength away from the open end of the microstrip line.

COMPUTER CODES

Five separate computer programs compose the programs written during this effort. They are briefly described below.

NAME	DESCRIPTION
1) speed2.f	Computes the MoM matrix elements for the waveguide to microstrip line transition. Has the fast feature of setting submatrix elements to zero if they represent an interaction physically separated by 25 or more strip

basis functions.

- 2) `sparameter.f` Using the output from `speed2.f`, this program calculates the S parameters for the symmetric microstrip-line-to-waveguide-to-microstrip-line transition. An even-odd mode analysis is used. The program adjusts the Y and T submatrices found in one, by taking into account waveguide fields that get reflected from a pec and a pmc termination alternatively.
- 3) `sparam2.f` Similar to `sparameter.f`, but it adjusts only the Y submatrix by adding the admittance of a pec and a pmc terminated waveguide alternatively. The results of this method appear to agree with those obtained by a program written by Marat.
- 4) `speedxwy.f` Similar to `speed2.f`, but this program includes y-dependent basis modes on the microstrip line.
- 5) `strip.f` Computes the MoM matrix elements for the stripline case. The basis modes on the stripline are x-directed and x-dependent.

Chapter 10. Conclusion

In summary and conclusion the following has been accomplished:

- 1) Comparisons of the results for the microstrip-waveguide-microstrip coupler analysis program with results from Marat's similar program are good.
- 2) Adding a y-dependence to the basis modes on the microstrip line brought the results closer to the experimental results. The experimental results are results of Marat and were compared by him in a "blind" test for us. We are not sure how much different these new results are from the experiment.
- 3) There remains one question about manipulating the result from half the problem (i.e. microstrip to waveguide) to the result for the whole problem (i.e. microstrip to waveguide to microstrip). Two methods of doing this disagree slightly in the final result of S_{11} versus frequency.
- 3) A program was developed and written for the stripline case. It seems to be working, but we have no experiment with which to test this.
- 4) The computer codes were developed on a SUN SPARC station 20, but are written in standard FORTRAN so are transportable.
- 5) The computer codes execution time was decreased tremendously, and more time can still be taken off.
- 6) Computer codes are included at the end of this report.
- 7) Parameter studies were done for the microstrip to waveguide to microstrip coupler. Some parameters tuned the frequency at which a match occurred, but as yet, no parameters were found to give a broadband (e.g. 4GHz to 6GHz) match.

References

- [1] D. M. Pozar, *Microwave Engineering*, Addison-Wesley Publishing Company, Inc., New York, 1990.
- [2] D. M. Pozar, "A Reciprocity Method of Analysis for Printed Slot and Slot-Coupled Microstrip Antennas," *IEEE Trans. Ant. and Propagat.*, vol. AP-34, December 1986, pp. 1439-1446.
- [3] Personal Communication with Marat Davidovitz.
- [4] F. J. Harackiewicz, "Circular Waveguide to Microstrip Line Transition," Final Report for Summer Faculty Research Program. Rome Laboratory, Hanscom AFB, MA, September 1994.
- [5] J. S. Bagby, C. H. Lee, Y. Yuan, and D. P. Nyquist, "Entire-Domain Basis MoM Analysis of Coupling Microstrip Transmission Lines," *IEEE Trans. Ant. And Propagat.*, vol. 40, no. 1, January 1992, pp. 49-57.
- [6] L. B. Felsen and N. Marcuvitz, *Radiation and Scattering of Waves*, IEEE Press, New York, 1994.
- [7] R. E. Collin *Field Theory of Guided Waves*, IEEE Press, New York, 1991.
- [8] M. Davidovitz and Y. T. Lo, "Rigorous Analysis of a Circular Patch Antenna Excited by a Microstrip Transmission Line," *IEEE Trans. Ant. And Propagat.*, vol. 37, no. 8, August 1989, pp. 949-958.
- [9] *Handbook of Microstrip Antennas*, edited by J. R. James & P. S. Hall. Peter Peregrinus Ltd., on behalf of the Institution of Electrical Engineers, London, UK, 1989.
- [10] W. S. Dorn and D. D. McCracken, *Numerical Methods with FORTRAN IV Case Studies*, John Wiley & Sons, Inc., 1972.
- [11] J. Pachner, *Handbook of Numerical Analysis Applications (with programs for engineers and scientists)*, McGraw-Hill Book Co., New York, 1984.
- [12] R. F. Harrington, *Time-Harmonic Electromagnetic Fields*, McGraw-Hill Inc., New York, 1961.
- [13] C. A. Balanis, *Antennas Theory Analysis and Design*, Harper & Row Publishers, New York, 1982.
- [14] M. Kobayashi and H. Sekine, "Closed-Form Expressions for the Current Distributions on Open Microstrip Lines," *IEEE Trans. Ant. And Propagat.*, vol. 39, no. 7, July 1991, pp. 1115-1119.

Appendix A Basis Function on the Strip with y Dependent Modes

The electric surface current on the strip is expanded in rooftop function with average of a few y-dependent modes:

$$\bar{J}_q(x, y) = \begin{cases} \left\{ 1 - \frac{|x - x_s - m_q h|}{h} \right\} \cdot \frac{T_{2i_q}(2y/w)}{\sqrt{1 - (2y/w)^2}} \hat{x} & \text{for } |x - x_s - m_q h| < h \text{ and } |y - y_s| < w/2 \\ 0 & \text{esle} \end{cases} \quad (\text{A-1})$$

where $q = 1, 2, 3, \dots, N$ (mode index for strip)

$m_q = 1, 2, 3, \dots, N_x$ (mode index for rooftop function)

$i_q = 1, 2, 3, \dots, N_y$ (mode index for y-dependent modes with edge conditions)

The Fourier transform of this current mode is

$$\begin{aligned} k_o \tilde{\bar{J}}_q &= k_o F_{x_{m_q}}(k_x) F_{y_{i_q}}^T(k_y) \hat{x} \\ k_o F_{x_{m_q}}(k_x) &= e^{jk_x x_s} e^{jk_x m_q h} h k_o \text{sinc}^2 \left(\frac{1}{2} \frac{k_x}{k_o} h k_o \right) \\ F_{y_{i_q}}^T(k_y) &= (-1)^{i_q} \frac{\pi w}{2} J_{2i_q}(k_y w / 2) \\ k_x &= k_t \cos \alpha, \quad k_y = k_t \sin \alpha \end{aligned} \quad (\text{A-2})$$

where J is a Bessel function.

Appendix B Method of Moments Equations with y Dependent Modes

As shown appendix D of [4], the MoM matrix equation showing units and dimensions is:

$$\underbrace{\begin{bmatrix} T^{AS} & \eta_0 \left(-Y^{AA} + \frac{\delta_{mn}}{Z_n} \right) \\ Z^{SS}/\eta_0 & -T^{SA} \end{bmatrix}}_{(M+N) \times (N+M)} \underbrace{\begin{bmatrix} \eta_0 I_n \\ V_n^r \end{bmatrix}}_{(N+M) \times 1} = \underbrace{\begin{bmatrix} \eta_0 \left(Y^{AA} + \frac{\delta_{mn}}{Z_n} \right) \\ T^{SA} \end{bmatrix}}_{(M+N) \times M} \underbrace{\begin{bmatrix} V_n^i \end{bmatrix}}_{M \times 1} \quad (B-1)$$

where M is the number of waveguide modes and N is the number of modes on strip

Here the definition of submatrix Y^{AA} remains the same as in Appendix D, but the definitions of submatrices T^{SA} , T^{AS} and Z^{SS} should be changed as follows:

The MxN matrix T:

$$T_{pq}^{AS} = \frac{C_{mp}}{4\pi^2} \int_0^\infty \left(\frac{k_t}{k_0} \right) d \left(\frac{k_t}{k_0} \right) \left[\tilde{b}_p(k_t) \cdot \hat{k}_t (hjte) I_{pq}^1(k_t) + \tilde{b}_p(k_t) \cdot \hat{z} \times \hat{k}_t (hjtm) I_{pq}^2(k_t) \right] \quad (B-2)$$

where $C_{mp} = 2\pi j^{(m-1)} \sqrt{2 - \delta_{0m}}$

$p = 1, 2, 3, \dots, M$ (mode index for waveguide)

$q = 1, 2, 3, \dots, N$ (mode index for strip)

and where

$$I_{pq}^1 = (-1)^{m_p-1} \int_0^{2\pi} \begin{pmatrix} 0^{TM} \\ c_{mp}^{TE}(\alpha) \end{pmatrix} F_x k_0 F_y^T \sin \alpha d\alpha$$

$$= \begin{cases} 0 & \text{if } p \text{ is TE and } \cos \phi \text{ harmonic and if } y_s = 0 \\ 0 & \text{if } p \text{ is TM} \\ 4 \int_0^{\pi/2} \text{Re} \left\{ F_{x_{mq}} k_0 \right\} F_{y_{iq}}^T \sin(m_p \alpha) \sin \alpha d\alpha & \text{if } p \text{ is TE and } \sin \phi \text{ harmonic and } m_p \text{ is odd} \\ -4j \int_0^{\pi/2} \text{Im} \left\{ F_{x_{mq}} k_0 \right\} F_{y_{iq}}^T \sin(m_p \alpha) \sin \alpha d\alpha & \text{if } p \text{ is TE and } \sin \phi \text{ harmonic and } m_p \text{ is even} \end{cases} \quad (B-3)$$

$$I_{pq}^2 = (-1)^{m_p-1} \int_0^{2\pi} \begin{pmatrix} C_{m_p}^{TM} \\ d_{m_p}^{TE} \end{pmatrix} F_x k_0 F_y^T \cos \alpha d\alpha$$

$$= \begin{cases} 0 & \text{if } p \text{ is TE and } \cos \phi \text{ harmonic} \\ 0 & \text{if } p \text{ is TM and } \sin \phi \text{ harmonic} \\ \pm 4 \int_0^{\pi/2} \operatorname{Re} \left\{ F_{x_{m_p}} k_0 \right\} F_{y_{i_q}}^T \cos(m_p \alpha) \cos \alpha d\alpha & \begin{cases} + \text{ if } p \text{ is TM, } \cos \phi \text{ harmonic and } m_p \text{ is odd} \\ - \text{ if } p \text{ is TE and } \sin \phi \text{ harmonic and } m_p \text{ is odd} \end{cases} \\ \mp 4j \int_0^{\pi/2} \operatorname{Im} \left\{ F_{x_{m_p}} k_0 \right\} F_{y_{i_q}}^T \cos(m_p \alpha) \cos \alpha d\alpha & \begin{cases} - \text{ if } p \text{ is TM and } \cos \phi \text{ harmonic and } m_p \text{ is even} \\ + \text{ if } p \text{ is TE and } \sin \phi \text{ harmonic and } m_p \text{ is even} \end{cases} \end{cases}$$

(B-4)

The two matrices T_{pq}^{SA} , T_{qp}^{AS} are related as in [4].

The NxN matrix Z:

$$Z_{pq}^{SS} / \eta_b = (-1)^{i_p + j_q} \frac{(hk_0 w k_0)^2}{4} \int_0^\infty \left(\frac{k_t}{k_0} \right) d \left(\frac{k_t}{k_0} \right)$$

$$\cdot \int_0^{\pi/2} d\alpha \left\{ \cos^2 \alpha (ejtm) + \sin^2 \alpha (ejte) \right\} \cos \left(k_0 h \frac{k_t}{k_0} \cos \alpha (m_p - n_q) \right) \operatorname{sinc}^4 \left(\frac{1}{2} \cos \alpha h k_0 \frac{k_t}{k_0} \right)$$

$$\cdot J_{2i_p} \left(\frac{k_t}{k_0} \sin \alpha \frac{k_0 w}{2} \right) J_{2j_q} \left(\frac{k_t}{k_0} \sin \alpha \frac{k_0 w}{2} \right)$$

(B-5)

where $p, q = 1, 2, 3, \dots, N$ (mode index for strip), $m_p, n_q = 1, 2, 3, \dots, N_x$ (mode index for rooftop function), and $i_p, j_q = 1, 2, 3, \dots, N_y$ (mode index for y-dependent modes with edge conditions).

Notice that the matrix Z^{SS} is Toeplitz by blocks here as the following:

$$[Z] = \begin{bmatrix} [z]_{11} & [z]_{12} & \cdots & [z]_{1N_x} \\ [z]_{21} & [z]_{22} & \cdots & [z]_{2N_x} \\ \vdots & \vdots & \ddots & \vdots \\ [z]_{N_y 1} & [z]_{N_y 2} & \cdots & [z]_{N_y N_x} \end{bmatrix} \begin{matrix} \updownarrow N_y \\ \updownarrow N_y \end{matrix} \quad (\text{H-6})$$

where $[z]_{m_p n_q} = [Z_{mn}]$ and $[Z]_{ij}$ has dimension $(N_x \times N_y) \times (N_x \times N_y)$. Thus, if one row of submatrices is known, the remaining submatrices may be filled by the algorithm

$$[z]_{m_p n_q} = [z]_{1, |m_p - n_q| + 1} \quad m_p \geq 2, \quad n_q \geq 1. \quad (\text{B-7})$$

Human Expert Identification of Latin American Dialects

Beth L. Losiewicz
Assistant Professor
Department of Psychology

The Colorado College
Colorado Springs, CO

Final Report for:
Summer Research Extension Program

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base
Washington, DC

and

Rome Laboratory

December 1995

HUMAN EXPERT IDENTIFICATION OF LATIN AMERICAN DIALECTS

Beth L. Losiewicz, Assistant Professor, Psychology Department
The Colorado College

ABSTRACT

The ability of human expert listeners to perform sub-language (Latin American Spanish) dialect identification was tested. Twelve dialectologists attempted to identify the country and city of origin of 90 Latin Americans, from spontaneous speech segments ranging in length from 2 to 30 seconds. On the average, dialectologists only correctly identified the speakers' country of origin 24% of the time. (Chance expectation was 9%). Chileans, Cubans, Puerto Ricans, Argentineans, and Colombians were identified best (in that order), and no cities were identified with greater than chance accuracy. There did not appear to be any strong systematic effect of segment length on identification accuracy, although the types of cues reported did vary with segment length. The dialectologists were not very confident in their decisions overall, although there was a significant correlation between their degree of accuracy and their confidence in their decisions. This correlation was highest for those dialectologists who performed best on the identification task.

When a country was not correctly identified, it was, more often than not, identified as another country of the same broad dialect region (i.e., Caribbean vs. Not-Caribbean). Puerto Rico/Cuba; and Venezuela/Puerto Rico were clearly mutually confusable, and there appeared to be a trend for Nicaragua/Venezuela; Peru/Colombia; Venezuela/Colombia; Chile/Argentina; Chile/Cuba; and Cuba/Venezuela to be mutually confusable.

Intonation and phonetic cues were most often reported by the dialectologists as the most important for correct country identification, with intonation reported slightly more often than phonetic cues. This advantage for intonation increased as the segments became shorter, and as they became more difficult to identify. Rhythm was the next most often reported cue, and also became relatively more important as the segments became shorter. Speed of speech, and lexical items apparently played a small role in dialect identification, and their importance did not appear to vary with identification difficulty nor segment length. Similar types of cues were reported for all eleven countries.

In conclusion, it was hypothesized that the sub-language identification performance of human experts is a complex task, relying on multiple constraining cues which can be used somewhat flexibly depending on task constraints. It was predicted that the best machine performance would be obtained with parallel algorithms exploiting both phonetic and intonational cues.

HUMAN EXPERT IDENTIFICATION OF LATIN AMERICAN DIALECTS

Beth L. Losiewicz, Ph.D.

OBJECTIVES

This study was conducted as a contribution to the overall research goal of developing techniques to subdivide languages based on information from the acoustic signal. The objectives were :

- To establish sub-language identification performance of trained human listeners
- To determine the degree of confidence trained human listeners have in their sub-language identification decisions
- To assess whether there is a correlation between decision confidence and identification accuracy
- To identify the properties of the acoustic signal on which trained human listeners base their identification decisions

INTRODUCTION

Speech segments of varying length, type and dialect were extracted from the Rome Labs/MIT Spanish Dialect Database. An analog audio tape was created from these segments, and Spanish dialectologists were asked to listen to the recorded speech segments and make a decision as to the dialect of origin of the speaker, choosing from a list of choices offered on a response sheet. They were also asked to rate their degree of confidence in their dialect decision, and indicate what types of cues led them to their decision. Responses were analyzed to achieve the experimental objectives outlined above.

SUBJECTS

Selection of Dialectologists

Trained dialectologists were recruited from throughout the United States via Linguist List (a linguistics electronic bulletin board), and 'word-of-mouth' within the Spanish linguistics community. As an incentive to participate, they were allowed to keep the test tape, and sent a list of the regional identity of the speakers on the test tape after they completed the identification task. A total of 28 dialectologists and Spanish linguists volunteered to participate. Dialectologists were mailed a copy of a prepared tape, and a response sheet, and asked to complete the test and return their responses within 30 days. Each dialectologist was also asked to complete a biographical data sheet, and submit a resume. (A sample of the biographical data sheet appears in Appendix C of the full technical report, available at Rome Labs, IRAA, Speech Processing, Rome, New York.) Responses from the 12 dialectologists who responded within the deadline were included in the detailed analyses below.¹

Selection of Speakers

Criteria One: Countries and cities represented:

- At least one sample from each city and country represented in the RL database was used.
- For countries/cities with few speech segments, all available segments were used (i.e., Venezuela, Chile, Argentina, Puerto Rico, Costa Rica, Mexico, Ecuador and Nicaragua)
- For countries with ample speech segments in the database (i.e., Peru, Cuba, Colombia), segments were chosen such that they were distributed across available cities

Criteria Two: Sex of Speaker

- In so far as possible (given limitation of segments available in the database and the city/country distributions) an equal number of male and female speakers were chosen.

¹Thanks to participants Hugh Buckingham, Felice Coles, John B. Dalbor, Nestor Ferrer, Alan M. Gordon, Jorge Guitart, Robert M. Hammond, Lee Hartman, Barbara A. Lafford, Fernando Martinez-Gil, Oscar H. Moreno, Bret Parker, Alfredo B. Torrejon, J. Diego Quesada, Kathleen Wheatley, Kirk A. Widdowson, and others, who asked not to be named.

Criteria Three: Segment Length

- Thirty Short, 30 Medium and 30 Long segments were extracted from the available speakers in the database. Each segment represents a unique speaker. Short segments were 2-3 seconds in duration, medium were 5-7 seconds, and long were 25-30 seconds. Available countries and cities of origin were as evenly distributed across the three types of segments as possible (i.e., within the limits of segments available, an equal number of each country/city appeared in each segment length).

Criteria Four: Choice of Segment from each Speaker

- The first easily extracted speech segment of the appropriate length for each speaker was chosen for the test segment. Segments were chosen regardless of content of speech, either in phonemes, words or semantic content; except that segments containing reference to place names or typical features of climate, geography or diet (that might have given a clue to the origin of the speaker) were not used.
- In as far as possible, segmentation was made at natural phrase or constituent boundaries. As a result, individual segments varied slightly in length within the limits for each Section length.

A summary of the distribution of length, country, city and sex of speakers in the selected segments is available with the full technical report at Rome Labs, IRAA, Speech Processing, Rome, New York.

PROCEDURES

The prepared audio tape was composed of three separate sections each containing 30 spontaneous speech segments of varying lengths, from a total of 90 different speakers. Section One contained segments of approximately 30 seconds. Section Two contained 5-7 second segments, each appearing twice in succession. Section Three contained 2-3 second segments, each of which appeared three times in succession. Ss were asked to listen to the speech segments, and circle the name of the country of origin of each speaker they heard. They were also asked to indicate their degree of confidence in each dialect decision on a scale of 0 (guessing) to 4 (very

confident), and to indicate what cues they used to make that decision. An example of the response sheet is available with the full technical report at Rome Labs, IRAA, Speech Processing, Rome, New York. The tape was approximately 70 minutes long, and the dialectologists were allowed to complete the task in two to three separate sittings, but were asked *not* to rewind the tape to re-listen to any segment(s).

Speech segments were randomized within each section, with all Long segments (30 speakers) appearing in Section 1, all Medium (30 speakers) in Section 2, and all Short (30 speakers) in Section 3. (Appendix B of the full technical report contains a list of the speakers included in each Section.)

Two different tapes were made with two different random orders within each Section (Tape A and Tape B), but on both tapes Section 1 (Long) appeared first, followed by Sections 2 (Medium) and 3 (Short).² Speech segments that had been previously identified (by the linguist who made the recording) as atypical, or those with an obvious American "accent" were not used.

On the tape, each speech segment was identified (in Spanish) both by its order in the taped sequence for each Section (e.g., "Segmento 1, Segmento 2," etc.) and by a unique speaker identification number derived from the Rome Labs database ID number for that speaker (e.g., "Voz 261"). Section 1 was preceded by extensive taped instructions (in Spanish) along with two practice speech segments. Sections 2 and 3 were likewise be preceded by two practice segments each. (The Spanish tapescript, and an English translation, appear in Appendix C of the full Technical Report.)

Speech segments were digitally copied to individual computer files using GW Instruments Soundscope16, with Digidesign Audiomedia II 16 bit board, at a sampling rate of 22Khz. Segments were then edited within Soundscope and (digitally) compiled onto master DAT tape (A and B) with appropriate instructions.³ A 3 second pause was inserted between the identification of the segment (by segment number and speaker ID number) and the segment itself. The two

²Section order was not counterbalanced, as it was assumed that the shortest segments would be very difficult to identify and might undermine the respondents' confidence if they appeared first.

³Equipment problems resulted in some of the segments on Tape A to be either speeded up or slowed down approximately 8%. This resulted in an apparent lowering (or raising) of the pitch of the voice, but there is no evidence that it affected dialect identification. This problem was remedied on Tape B.

repeats of each segment in Section Two, and the three repetitions in Section Three were separated by 3 seconds of silence. Fifteen seconds of silence separated each new segment.

ANALYSIS OF RESULTS

OVERALL ACCURACY: Country, City, and Caribbean vs. Non-Caribbean

The identification task was apparently a very difficult one. On the average, any given dialectologists only identified 15% of the speakers, although the range of performance was fairly large: from 4% correct for the least accurate dialectologist in his least accurate section, to 50% for the most accurate in his most accurate section. This performance was significantly better than chance, $t(10) = 3.15$, $p < .01$.⁴

Of the 90 speakers, 82% were correctly identified (by country) by at least one of the twelve dialectologists, while 16 were never identified correctly. (Appendix A, available in the full technical report at Rome Labs, Speech Processing, contains a list of all speakers in rank order of their (country) identifiability by the respondents.)

Accuracy for city identification was also rather poorer than expected. Overall, only 12% of the time were respondents able to identify the city of origin of those speakers for whom they correctly identified the country, and that only if the correct city was provided as one of the choices on the answer sheet. Except for Colombia, which had more than one choice of city, chance alone would predict either a 50% or 33% accuracy rate (depending on whether "unknown" and "other" were considered as two separate responses, or lumped together in the category "wrong response"). Presumably this below-chance performance resulted from the dialectologists being conservative in their identification responses, choosing "unknown" or "other" whenever they were not certain. Also, some respondents never indicated a choice of city, or did so only sporadically: another factor which contributed to the below chance accuracy rate for city identification.

⁴Approximately 9% accuracy (1/11 countries) would be expected by chance alone. T-test was performed on adjusted percent correct by country. See Appendix A "Calculations" for the raw data.

The majority of the respondents included some note on their response sheets expressing their dismay at the difficulty of the identification task, often adding that this was the result of the lack of specific (phonetic) cues on which to base a decision. A number also noted that, even when cues present, they were insufficient to identify a specific country, and were only useful for making rough divisions (e.g., coastal vs. highland).

The poor identification performance is probably due, in part, to selection biases inherent in the database. First, all the speakers to be identified were, at least temporarily, living in the United States at the time of the recording interview. Second, it is likely that they were socio-economically better off and better educated⁵, on the whole, than a representative sample collected in the actual country of origin. By the very fact of their presence in Miami they are known to have traveled away from home⁶, probably more than the norm. They were all literate (they either responded to a Spanish newspaper advertisement, or were recruited on a University campus). Lastly, although care was taken to make the informants comfortable, and to encourage them to speak informally and spontaneously during the interview, the unfamiliar environment, presence of strangers, and the microphone were likely to inhibit the informal, spontaneous speech most likely to exhibit dialect characteristics. As a result of these inherent selection biases, very great caution must be exercised in generalizing the results of this study to any population that is not reasonably well-off, well-educated, and possibly well traveled, in a relatively formal environment.

Table 1 summarizes accuracy by city and country, both overall and for Caribbean vs. non-Caribbean countries. Also included are results of an additional analysis which investigated how well respondents could correctly identify general dialect region (i.e., Caribbean (or Coastal) vs. Non-Caribbean (or Highland) whether or not they correctly identified the correct country.⁷ The Table contains both raw percentage of correct identifications, and (probably a more useful)

⁵Education level ranged from "Some High School" to "Some Graduate Education." The average speaker had completed high school and at least some college. See Appendix B for education data for each speaker.

⁶Important both because dialect features are known to shift with contact with speakers of other dialects, and because strength of identification with, and plans to stay in, a particular region are known to affect strength of dialect characteristics in an individual's speech.

⁷For the purpose of this analysis, **Caribbean** was defined as: Cuba, Puerto Rico, Venezuela, Nicaragua, Costa Rica and Caribbean Coastal Colombia. **Not Caribbean** was defined as: Mexico, Peru, Ecuador, Non-Caribbean Colombia, Chile and Argentina. This division was chosen in consultation with several dialectologists and is not beyond dispute. For example, Costa Rica might be classified as Caribbean or Not Caribbean in terms of dialect features. The choice was made to include it as "Caribbean" both for instrumental and *post-hoc* reasons: that choice made it easier to calculate results by having an equal number of countries in each group, and, in the majority of mis-identifications, it was mis-identified as a Caribbean country.

adjusted percentage, from which the number of correct identifications expected by chance alone has been subtracted. (This adjusted figure was particularly necessary for the Caribbean/Non-Caribbean division, as the total number of speakers from each category in the test segments was uneven (Caribbean = 28 speakers; Non-Caribbean = 62 speakers)). (The probability calculations used to adjust these figures are included in Appendix A in the full technical report.).

TABLE 1:

SUMMARY OF PERCENTAGE OF CORRECT ANSWERS

(Raw percentages, and Percentages adjusted for Chance)

	Raw % By Country	Adjusted % By Country*	Raw % By City	Adjusted % By City*	Raw % Caribbean vs. Non-Caribbean	Adjusted % Caribbean vs. Non-Caribbean*
Not Caribbean	23	14	14	0	59	9
Caribbean	26	17	12	0	69	19
ALL	24	15	13	0	62	12

The division of the data into Caribbean vs. Non-Caribbean speakers allowed an analysis of differential identification accuracy rates between the two dialect divisions. (Although there were fewer speakers in the Caribbean category, they were sufficient in number to believe that the adjusted figures in Table 1 are reasonably accurate estimates of the dialect population as a whole.) First, it should be noted that both country and city identification were better for Non-Caribbean than for Caribbean countries. However, an interesting reversal occurs when the analysis is changed to only ascertain whether the respondents were able to accurately place each speaker in either the Caribbean or Non-Caribbean category. In that analysis, respondents were better at correctly identifying a Caribbean speaker as being Caribbean than a Non-Caribbean speaker as being Non-Caribbean. This pattern of results could be indicative of a greater homogeneity and (as a group) salience of dialect features *within* the Caribbean dialect group, and both less

homogeneity, and less overall salience of dialect features in Non-Caribbean dialects. This would allow more accurate country-by-country distinctions in the latter group, and more accurate overall identification in the former.

PERCENTAGE ACCURACY BY COUNTRY

Table 2 summarizes identification accuracy country by country.⁸ As in Table 1, these figures are expressed in percentage of correct identifications, adjusted for percentage correct expected by chance alone.⁹ Information about the number of speakers from each country is also included, since this information bears directly on the generalizability of these results to the whole population of that country. In short, the more speakers from a given country, the more accurately the numbers in Table 2 should reflect the identifiability of the entire population of speakers from that country. Thus special caution is advised in interpreting the accuracy ratings for those countries with only one speaker in the test segments: Mexico, Ecuador, Costa Rica, Nicaragua.¹⁰

Leaving aside those countries represented by only a single speaker, it appears that, of all the Non-Caribbean countries, Chile is the most identifiable (35%), with Argentina and Colombia second (14% and 12%), and Peru lagging far behind (5%). In the case of Chile, there is good reason to believe that the high percentage of accuracy is truly a reflection of very distinctive dialect features for this country.¹¹ The exact cause of the low performance for Peruvians is less clear. It could be the result of fewer salient dialect features in Peruvian Spanish (an interpretation suggested by the fact that at least one dialectologist specifically stated in a note that he (correctly) chose Peru as the country of origin because of *lack of* distinctive (non-standard) dialect features). It is also possible that Peruvian Spanish was less well identified because it has been studied less, or because dialectologists have less day-to-day exposure to Peruvians.¹²

⁸The two speakers from Barranquilla (Coastal Colombia) are analyzed separately from the other Colombian speakers in this Table.

⁹Again, details about the probability calculations on which the adjustments were based are included in Appendix A.

¹⁰Just as a single toss of a penny will, necessarily, give an over- or under-estimate of how many times "heads" will appear in a series of tosses (i.e., 100% or 0%), so estimates based on the identifiability of a single speaker probably severely over- or under-estimate the true identifiability of the population as a whole. It is not surprising that the highest, and lowest identification accuracies are from these countries with very small sample sizes.

¹¹Although it is interesting to point out that not even the dialectologist/respondent who was a native speaker of Chilean Spanish was able to correctly identify *all* the Chilean speakers in the test segments.

¹²That is, perhaps fewer Peruvians than Chileans, Argentineans or Colombians reside in the U.S., resulting in less personal familiarity with speakers of that dialect by the dialectologists. One respondent suggested that the best

TABLE 2

PERCENTAGE CORRECT IDENTIFICATION OF EACH COUNTRY
(Adjusted for Chance)

NUMBER OF SPEAKERS IN TEST POOL	COUNTRY	COUNTRY: ADJUSTED* % CORRECTLY IDENTIFIED	ADJUSTED* % CORRECTLY IDENTIFIED AS CARIBBEAN VS. NOT CARIBBEAN:
1	Mexico	49	8
8	Chile	35	13
9	Argentina	14	5
5	Colombia	12	10
10	<i>Bogota</i>	<i>13</i>	8
5	<i>Medellin</i>	<i>13</i>	8
8	<i>Cali</i>	<i>9</i>	<i>10</i>
5	<i>Other (Non-Caribbean)</i>	<i>13</i>	<i>13</i>
	<i>Col</i>		
15	Peru	5	11
1	Ecuador	0	8
13	Cuba	25	22
7	Puerto Rico	22	23
4	Venezuela	4	0
2	<i>Coastal Col.</i>	<i>0</i>	<i>25</i>
1	Costa Rica	0	42
1	Nicaragua	0	8

* Percent correct not accounted for by chance alone.

identification performance would likely be for those countries represented on the Spanish *Novelas* (Soap Operas), because respondents would be more personally familiar with those dialects. This hypothesis was not investigated in detail, although, based on the *Novelas* available in Colorado, Venezuela would be the best identified if this were true.

Of the Caribbean countries in test sample, Cubans were most accurately identified (25%), followed closely by Puerto Ricans (22%). Venezuela finished a poor third (4%), either, again, because that country lacks salient dialect cues or because the Venezuelan dialect is less studied and/or Venezuelan speakers less common in the U.S.

TABLE 3

RAW PERCENTAGE CORRECT IDENTIFICATION OF CITIES
(50% identification was expected by chance alone)

NUMBER OF SPEAKERS IN TEST POOL	CITY*	RAW % CORRECTLY IDENTIFIED
1	Guadalajara, Mexico	33%
7	Santiago, Chile	29%
5	Buenos Aires, Argentina	18%
21	Colombia Overall	6%
9	Bogota, Colombia	12%
5	Medellin, Colombia	0%
7	Cali, Colombia	1%
9	Lima, Peru	17%
1	Quito Ecuador	8%
9	Havana, Cuba	15%
1	Caracas, Venezuela	0%

*Managua, Nicaragua, San Jose, Costa Rica and Quito Ecuador and San Juan, Puerto Rico are not included because no speaker was correctly identified as being from one of these countries.

PERCENTAGE ACCURACY BY CITY

Although *no* city was correctly identified with greater than chance accuracy, for comparison purposes raw accuracy figures (given that the country was correctly identified) are summarized in Table 3. Santiago, Chile was by far identified the most often (raw accuracy = 29%), followed by Buenos Aires, Argentina (18%), Lima, Peru (17%), and Havana Cuba (15%).¹³

It is also interesting to note that, in spite of common agreement among dialectologists that the dialects of the major Colombian cities (e.g., Bogota, Medellin, Cali) are distinctive, in this study they were not very distinguishable (other than, possibly, distinguishing between Caribbean vs. Highland regions).

ACCURACY BROKEN DOWN BY LENGTH OF SEGMENT

The respondents' accuracy of country identification also varied by the length of the speech segment. Surprisingly, though, the least accuracy was obtained for the *middle* length of segment (5-7 seconds), not the shortest. The most likely explanation would appear to be that speakers in that section were simply more difficult to identify, independent of segment length. Mean accuracy rates are in Table 4.

TABLE 4

MEAN ACCURACY RATES BY LENGTH OF SEGMENT

(Not Adjusted for Chance)

Length of Segment	MEAN ACCURACY
30 seconds (Long)	27%
5 -7 seconds (Medium)	20%
2 - 3 Seconds (short)	24%

¹³Even these accuracy rates might have resulted not from unique and salient dialect features for the city given, but from the relative political and cultural salience of those cities - that is, it may be more likely in the U.S. to meet a Chilean from Santiago than otherwise; whereas Colombians in the U.S. may be more equally distributed across different cities. If so, a dialectologist who does not specialize in Chilean dialects may be more likely to recognize a Chilean speaker as being from Santiago, or more likely to label all Chileans as Santiagan.

DEGREE OF CONFIDENCE:

Overall the dialectologists were not very confident in their choices. The mean confidence rating was 2.2 (SD .63; range 1.07 - 3.03) on a 1 to 4 scale (1= not at all confident and 4 = very confident). This lack of confidence was also reflected in the written comments of many of the participants.

TABLE 5:

CORRELATION BETWEEN ACCURACY AND CONFIDENCE: (Not adjusted for Chance)

Length of Segments	Overall (n=360)	Three Most Accurate Dialectologists (n=90)	Three Least Accurate Dialectologists (n=90)
30 seconds	.33**	.35**	.21*
5-7 seconds)	.15**	.27**	.02 ns
2-3 seconds	.27**	.50**	.31**

** p< .01

*p < .05

ns = not significant

Table 5 summarizes the correlations between confidence level and accuracy. There was small but significant correlation between each dialectologists' mean confidence level and that dialectologists' accuracy for each section ($r(34) = .45, p < .01$).¹⁴ Interestingly the degree to which their confidence level predicted their accuracy varied between Sections, and varied between Dialectologists as seen in Table 5. The three dialectologists who most accurately identified the

¹⁴The correlation was computed on the mean confidence score (1 - 4) and the mean accuracy for each Section for each Dialectologist. N = 36 (3 sections x 12 Dialectologists).

speech segments also were most accurate in assessing their own accuracy. The segments with the least overall accuracy (5-7 seconds) also have the smallest correlation between degree of confidence and accuracy. Apparently greater accuracy implies better ability to judge that accuracy.¹⁵

MIS-IDENTIFICATIONS

Table 6 is a confusion matrix of the mis-identifications made by the respondents. Chance alone would predict that any given country would be mis-identified approximately 9% of the time as any other given country. In the matrix, those confusions that exceeded chance expectations are in bold print. Considering only those mis-identifications that exceeded chance, the only clear pattern that is apparent is that, on the whole, Caribbean countries are apt to be mis-identified as another Caribbean country, and Non-Caribbean countries are apt to be mis-identified as another non-Caribbean country. Surprisingly, there appears to be very little symmetry in the mis-identifications. For example, although a Costa Rican was mis-identified as a Puerto Rican 67% of the time, no Puerto Rican was ever mistaken for a Costa Rican. A certain portion of this asymmetry arises from the unequal representation of the different countries in the test samples. (For example, there was only one speaker from Costa Rica, which, as, discussed earlier, makes the statistics for Costa Rica rather unreliable). But it does not seem likely that the lack of mutual mis-identifications can be wholly ascribed to this unequal representation.

Table 6 summarizes *all* the pairs of mutual mis-identifications for which *any* mis-identifications occurred - even if those mis-identifications were well below that expected by chance. (NB: Extra care must be taken in interpreting results for those countries represented by only one speaker (marked with an asterisk in the Table).) Although there still appears to be a great deal of "noise" in the data, it seems fairly clear from this Table that Puerto Rico/Cuba, and probably Venezuela/Puerto Rico are mutually confusable. Less certain, but still probable, is the trend for mutual confusability between Nicaragua/Venezuela; Peru/Colombia; Venezuela/Colombia; Chile/Argentina; Chile/Cuba; and Cuba/Venezuela. Again, with the salient exception of Venezuela, only Caribbean countries are confused with other Caribbean countries, and only Non-Caribbean with other Non-Caribbean.

¹⁵This is in line with recent research on the nature of expertise: the better one is at a task, the better one is able to accurately judge one's own performance on that task.

TABLE 6: PERCENTAGES OF MIS-IDENTIFICATION PAIRS (Above chance percentages are in bold print)

COUNTRY OF ORIGIN												
		NOT CARIBE						CARIBE				
		Chil	Arg	Col	Mex	Ecu	Peru	CR	Nic	Ven	PR	Cuba
MIS- O A	N C		7%	1%			8%			8%	6%	4%
	Chile											
	Arg	10%		2%			7%		8%			1%
	Col	3%	2%			17%	13%		17%	10%	2%	5%
IDENTI- I B	Peru	3%	8%	8%		17%				8%		5%
	Mex	2%	7%	20%		17%	14%			19%	1%	2%
	Ecu		7%	6%			4%		8%	2%	5%	3%
FIELD AS C	CR	3%	5%	7%	25%		3%		8%	4%		2%
	Nic	2%	6%	8%		8%	5%			15%	8%	3%
	Ven	8%	8%	7%			6%		17%		12%	12%
	PR	7%	8%	6%		8%	7%	67%	8%	13%		21%
	Cuba	10%	9%	5%			5%	17%	25%	2%	21%	

TABLE 7

MIS-IDENTIFICATION PAIRS (9% mis-identifications expected by chance alone)

	First Country Mis- Identified as Second (Percentage)	Second Country Mis- Identified as First (Percentage)
Both Countries Mis-Identified as the Other With Greater Than Chance Frequency		
PR / CUBA	21	21
VEN / PR	13	12
NIC* / VEN	17	15
Only One of the Two Countries Mis-Identified With Greater Than Chance Frequency		
PERU / COL	13	8
VEN / COL	10	7
CHILE / ARG	10	7
CHILE / CUBA	10	4
CUBA / VEN	12	2
ARG / CUBA	9	1
NIC* / CUBA	25	3
ECU* / COL	17	6
ECU* / PERU	17	4
NIC* / COL	17	8
C RICA* / CUBA	17	2
Neither Country Alone Mis-Identified With Greater Than Chance Frequency		
VEN / CHILE	8	8
ARG / PERU	8	7
PERU / CHILE	8	3
CHILE / COL	3	1
VEN / PERU	8	6
CHILE / PR	7	6
PERU / CUBA	5	5
COL / CUBA	5	5
COL / PR	6	2
ARG / COL	2	2
NIC* / ECU	8	8
NIC* / PR	8	8
ECU* / PR	8	5
NIC* / ARG	8	6

* Only one speaker from this country in test segments

Bold print = those pairs judged to be mutually confusable (above chance)

The lack of clear patterns of mutual confusions suggests that the profiles of the different dialects will not fall readily into neat, mutually exclusive (classical logic) categories. Rather, they seem to bear "family resemblances" to each other, and a more complex "prototype" approach to categorization will probably be needed to adequately characterize their similarities and differences.

CUE ANALYSIS: Overall

The response sheets, on which the dialectologists identified the speakers, also contained a list of possible dialect cues that could have been used in each identification. Dialectologists were asked to indicate the cues they used, and to indicate their relative importance in making each dialect identification. Caution must be used in interpreting this reported cue information, however. It is fairly well established that in speech perceptions, as in a variety of other perceptual and cognitive tasks, decision processes are not always open to conscious inspection. In fact, the cues people *think* they use to make a decision are often unrelated to the cues they actually *do* use. In light of this, it is very important to emphasize that the present data can only tell us about the cues the dialectologists *thought* they used, which may or may not reflect the cues they actually used, to make their dialect identifications.

Table 8 summarizes the relative frequency of report for each cue. Listed there is the total number of times each type of cue was reported for a correct identification, and a weighted score which take into account any rank ordering indicated by the dialectologists.¹⁶ These numbers were then converted to percentages to allow comparison between cues.

It is apparent from Table 8 that intonation was the most reported cue, followed closely by phonetic cues. The third most reported cue was rhythm, followed by speed of speech, and lexical cues. Voice pitch was almost never reported as a relevant cue.

¹⁶The Weighted Scores were computed by giving first ranked cues a value of 3 points; second ranked cues a value of 2; and third and fourth ranked cues a value of 1. Unranked cues were given a value of 3 points.

TABLE 8

CUES REPORTED FOR CORRECT IDENTIFICATIONS OF COUNTRY OF ORIGIN

	Total Number of Times Indicated	Weighted Score	Cue as % of Total Reported Cues	Cue as % of Weighted Score
Intonation	175	509	33%	35%
Phonemes	163	452	31%	35%
Rhythm	88	227	17%	16%
Speed	51	137	10%	9%
Lexical Items	45	120	9%	8%
Pitch	5	7	1%	.5%

CUE ANALYSIS: By Country

It is also of interest to know whether the respondents relied on different sorts of cues for the correct identification of different countries. Table 8 gives the weighted percentage of cues used, broken down by the country of origin of the correctly identified speaker.

This country by country cue analysis yields the same basic pattern seen in the data set as a whole, with a few exceptions. Mexico and Puerto Rico appeared to have a small trend for phonetic rather than intonation cues to be most important; Mexico and Argentina had lower than average use of speed as a cue; and Venezuela shows a lower than average report of the use of rhythm cues. All of these differences are very small, however, and may well either be "noise" in the data, or have other explanations. (For example, the elevated importance of the lexicon in the identification of the (single) Mexican speaker probably reflects the use of the phrase "*no mas*" in that segment, and is probably not generalizable to Mexican speech in general.).

In summary, we may reasonably conclude that there are no major differences in the cues reported for the correct identification of the different Latin American countries.

TABLE 9

CUE ANALYSIS BY COUNTRY (Weighted Percentages)*

	Mean	Arg	Chile	Col	Peru	Mex	PR	Cuba	Ven
Intonation	35	40	37	38	36	29	28	33	33
	31	37	36	28	19	34	36	32	25
Phonemes									
Rhythm	16	17	14	19	17	14	14	14	8
Speed	9	4	7	10	15	0	13	11	16
	8	2	4	6	13	23	8	10	16
Lexicon									
Pitch	.5	0	1	1	0	0	0	0	0

*The Nicaragua and Costa Rican speakers were never correctly identified, and the Nicaraguan speaker was only identified by one dialectologist, therefore these countries are not included in this analysis.

CUE ANALYSIS: By Difficulty of Speaker Identification

We can also ask whether the type of cue used was different for those speakers who were easy to identify and those who were hard to identify. Reported cues for speakers who were correctly identified by 5 to 11 of the twelve dialectologists (N=18) were analyzed separately as "High Identifiable Speakers." Speakers identified by at least 3 to 4 of the dialectologists (N=26) were analyzed as "Medium Identifiable Speakers," and those who were identified by only 1 or 2 dialectologists were analyzed as "Low Identifiable Speakers" (N=30). The results of this analysis by difficulty of speaker identification are summarized in Table 10.

Table 10

CUE ANALYSIS BY DIFFICULTY OF SPEAKER IDENTIFICATION

	HIGH IDENTIFIABLE		MEDIUM IDENTIFIABLE		LOW IDENTIFIABLE	
	Total %	Weighted %	Total %	Weighted %	Total %	Weighted %
Intonation	30%	32%	33%	35%	43%	45%
	33%	34%	32%	32%	21%	22%
Phonemes						
Rhythm	16%	14%	17%	16%	19%	18%
Speed	10%	10%	11%	11%	6%	6%
Lexicon	10%	10%	5%	5%	11%	9%
Pitch	1%	1%	1%	<.5%	0%	0%

The analysis in Table 10 suggests that, although intonation and phoneme identifiability appear to play equal roles for the relatively easier identifications, as the task becomes more difficult intonation cues play a larger role, at the expense of phonetic cues. A slight rise in report of rhythm as a cue also occurs with the more difficult identifications. This could result from one of two different cause. It is possible that as the identification task became more difficult, the respondents relied more on intonation and rhythm cues (or at least reported that they did). A second possibility is that a certain subset of dialectologists, those who are best at making identifications, are also more apt to use intonation and rhythm cues. Further analysis of the data (beyond the scope of the present paper) should allow us to chose between these alternative possibilities.

CUE ANALYSIS: By Length of Speech Segment

The last question is whether the types of cues used changed as a function of the length of the speech segments being identified. Table 10 summarizes the cues reported useful for the correct identification of Long (30 sec) , Medium (5-7 sec), and Short (2-3 sec) speech segments.

TABLE 11

CUE ANALYSIS, BY LENGTH OF SPEECH SEGMENTS (Weighted Percentages)

	LENGTH OF SEGMENTS			
		SHORT	MEDIUM	LONG
	Mean	(2-3 sec)	(5-7 sec)	(30 sec)
Intonation	35	39%	38%	30%
Phonemes	31	25%	30%	37%
Rhythm	16	19%	17%	12%
Speed	9	9%	11%	9%
Lexicon	8	8%	4%	11%
Pitch	.5	.2%	.5%	.7%
MEAN % CORRECT		24%	20%	27%

The report of both intonation and rhythm cues increase as the segments get shorter.¹⁷ It also suggests that the intonation cues being used are *not* supra-segmental cues (e.g., question intonation contours) as these would presumably only be evident in stretches of speech longer than 3 seconds.

CUE ANALYSIS: Summary

On the whole intonation cues and phonetic cues share honors as the most often reported dialect cues for correctly identified speakers, with intonation having a slight advantage. This advantage increases as the speech segments become shorter, and as they become harder to identify. It is possible either that intonation cues become more useful as segments get longer

¹⁷ It is unlikely that this pattern merely reflects the prior finding that, as the speakers become less identifiable, intonation becomes more important as a cue. Segment length and identifiability are apparently unrelated. As indicated earlier, overall accuracy does not decrease linearly with segment length. A closer look at the composition of the three identifiability groups corroborates this: 5, 12 and 7 *short* segments fell in the high, medium and low identifiable groups respectively; 5, 8, and 8 *medium* segments fell in those three groups; and 8, 6 and 6 *long* segments. No indication that length is directly correlated with identifiability is found here.

and/or harder, or that the dialectologists who are best at the identification task are more apt to rely on intonation cues either instead of, or in addition to, phonetic cues.

Rhythm is the next most important cue, and like intonation is reported slightly more often as the speech segments become shorter. Speed and lexicon are reported relatively infrequently, and their relative importance is apparently impervious to segment length.¹⁸

Although there is not enough evidence to draw firm conclusions, the interaction of the importance of phonetic, intonation and rhythm cues with the difficulty of identification and segment length suggests that the identification process, as performed by human experts, relies on multiple cues, the importance of which can be adjusted depending on task and circumstances. This is in line with most recent research on the nature of human cognition and speech perception: we take into consideration all available information and come to "best fit" decision. If any given piece or type of information is missing, we can still make a decision based on the remaining cues.

SUMMARY AND CONCLUSIONS

In brief, human expert identification of the countries of origin of a sample of fairly educated speakers from a variety of Latin American countries, based on short taped speech segments, is fairly poor, albeit better than chance. Length of segment (ranging from 2 to 30 seconds) does not seem to have a linear effect on identification accuracy, although it does appear to influence the type of cue used to make the identification.

The dialectologists who participated in the study were not particularly confident in the decisions they made, although their confidence ratings were positively correlated with accuracy. Further, the dialectologists who made the most accurate identifications had the highest correlation between degree of confidence and accuracy.

Some patterns emerged among the mistaken identifications. Overall, Caribbean countries were most often confused with other Caribbean countries, and Non-Caribbean countries were most often confused with other Non-Caribbean countries (with the salient exception of Venezuela which was often confused both with other Caribbean, and with Non-Caribbean countries). The most confusable pairs were Puerto Rico/Cuba, and Venezuela/Puerto Rico, and a trend for mutual confusability between Nicaragua/Venezuela; Peru/Colombia; Venezuela/Colombia; Chile/Argentina; Chile/Cuba; and Cuba/Venezuela was also evident.

¹⁸For lexical cues this is probably because the difference between 30 seconds and 3 seconds of speech is not enough to significantly change the probability that a dialect-unique lexical item will appear. No doubt lexical cues would become more important as segment lengths differences are greater - say 30 seconds versus 5 minutes.

The types of cues deemed important for dialect identification appear to be the same across the different Latin American dialects studied, although length of segment and the difficulty of the identification task appeared to have an effect on the type of cue used. The cues reported most often were intonation and phonetic cues, followed by rhythm. Speed and lexical cues were reported, but less often. As the segments became shorter, and as they become more difficult, the relative proportion of intonation and rhythm cues reported increased at the expense of phonetic cues while the proportion of speed and lexical cues reported remained stable.

The results of this study suggest that, although sub-language identification may be feasible, it is a relatively difficult task, even for human experts, and probably depends on the use of multiple types of acoustic cues. The current computer algorithms that concentrate on the exploitation of phonetic cues may well be insufficient to do accurate sub-language identification, especially on very short samples of speech. Intonation cues may be a better choice on which to build identification algorithms, and the best performance of all could probably be expected from parallel algorithms concurrently exploiting both phonetic and intonational cues.

Mohamed Musavi report unavailable at time of publication.

INFRARED IMAGES OF ELECTROMAGNETIC FIELDS

John D. Norgard
Professor/ECE
Director/Electromagnetics Laboratory
Department of Electrical & Computer Engineering

University of Colorado
Department of Electrical & Computer Engineering
College of Engineering & Applied Science
1420 Austin Bluffs Parkway
Colorado Springs, CO 80917-7150

Final Report for:
Summer Research Extension Program
Rome Laboratory

Sponsored by:
Air Force Office of Scientific Research
Bolling AFB, Washington DC
and
University of Colorado at Colorado Springs

December 1995

INFRARED IMAGES OF ELECTROMAGNETIC FIELDS

John D. Norgard
Professor/ECE
Director/Electromagnetics Laboratory
Department of Electrical & Computer Engineering
University of Colorado

Abstract

The unique measurement capabilities of the infrared (IR) imaging technique for mapping electromagnetic (EM) fields were further developed and enhanced during this extension period. The IR technique was used to map two-dimensional EM field distributions near radiating and scattering bodies. Initial tests to prove the validity, accuracy and sensitivity of this technique were performed this summer in the anechoic chamber at the Electromagnetic Vulnerability Analysis Facility (EMVAF) at Rome Laboratory (RL). Empirical calibration tests were performed during the extension period to determine the absolute intensity of the magnitude of the electric field. Initial microwave holography tests were performed to determine the feasibility of measuring the phase of the electric field. Several simple, canonical objects were tested to validate the accuracy of the IR technique.

i) Radiation tests were performed on a conducting cylinder irradiated by an incident plane wave. E- and H- field patterns of the scattered/diffracted energy from the cylinder are being used to validate a hybrid FEM/MoM Code which is being developed at JPL.

ii) Coupling tests were performed to determine the induced energy coupled through long, thin slot apertures at various orientations in the side of a cylinder. Demonstrations of these tests were performed at the International EMC Symposium in Atlanta.

iii) Scattering characteristics of several dielectric cylinders and spheres were also determined. Metallic and dielectric half-cylinders and half-spheres were tested for their scattering characteristics.

iv) Predicted (and measured) electric fields radiated from a standard gain horn antenna were used to calibrate the measured temperature levels in the IR detector screen.

v) The scattered electric field from a silver coated scale model F16 aircraft was mapped . The electric field intensity in the horizontal and vertical longitudinal planes through the fuselage of the aircraft are being used to validate the accuracy of the GEMACS numerical code.

vi) The phase of a large phased array antenna was measured.

Four papers were presented at international IR and EM conferences last summer; eight additional papers were presented during the extension period. One paper was published. Two seminars also were presented on the technique. Seven other papers have been submitted for presentation at conferences.

INFRARED IMAGES OF ELECTROMAGNETIC FIELDS

John D. Norgard

Professor/ECE

Director/Electromagnetics Laboratory

Department of Electrical & Computer Engineering

University of Colorado

Table of Contents

1. Introduction
2. IR Measurement Technique (Brief Overview)
 - 2.1 IR Experimental Setup
 - 2.1.1 Electromagnetic Parameters
 - 2.1.2 Thermal Parameters
 - 2.1.3 Thermal Equilibrium
 - 2.2 Approximate Solution
 - 2.3 IR Detector Screen
 - 2.3.1 Electric Field Detector Screen
 - 2.3.2 Magnetic Field Detector Screen
 - 2.4 IR Camera
 - 2.5 IR Images
 - 2.5.1 Spatial Resolution
 - 2.5.2 Thermal Resolution
 - 2.5.3 Thermogram Errors
 - 2.5.3.1 Lateral Conduction Effects
 - 2.5.3.2 Lateral Convection Effects
 - 2.5.4 IR Measurement Accuracy
 - 2.6 IR Advantages and Disadvantages
3. IR Thermograms
 - 3.1 Magnitude Estimates
 - 3.1.1 Relative Measurements

3.1.2	Absolute Measurements
3.1.2.1	Theoretical Calibration
3.1.2.2	Experimental Calibration
3.1.2.2.1	Experimental Setup
3.1.2.2.2	Experimental Data
3.2	Phase Estimates
3.2.1	Energy Minimization Technique
3.2.2	Microwave Holography Technique
3.2.2.1	Plane-to-Plane (PTP) Phase Retrieval Technique
3.2.2.2	IR Hologram Phase Retrieval Technique
4.	Results
4.1	Scattering from a Cylinder
4.2	Scattering from an F16 Scale Model Aircraft
4.3	Phased Array Antenna Patterns and Element Efficiency
5.	Summary of Extension Period Tasks
6.	Conclusions
7.	Future Work
8.	Publications
	Acknowledgments
	Appendix

INFRARED IMAGES OF ELECTROMAGNETIC FIELDS

John D. Norgard

Professor/ECE

Director/Electromagnetics Laboratory

Department of Electrical & Computer Engineering

University of Colorado

An infrared (IR) measurement technique is being developed to measure electromagnetic (EM) fields. This technique uses a minimally perturbing, thin, planar IR detection screen to produce a thermal map of the intensity of the EM energy deposited over a two-dimensional region of the screen. EM fields near radiating microwave sources and scattering bodies were measured with this technique. This technique also was used to correlate theoretical data with experimental observations and to experimentally validate complicated numerical codes which predict electric field distributions inside waveguide cavities (E-Fields) and surface current distributions on metallic surfaces (H-Fields).

1. Introduction

A non-destructive, minimally perturbing IR measurement technique is being developed to observe EM fields. Metallic surface currents and charges also can be measured with this technique.

This IR measurement technique produces a two-dimensional IR thermogram of the intensity of the electric or magnetic field being measured, i.e. a two-dimensional isothermal contour map or a gray scale of the intensity of the EM field.

Electric and magnetic fields have been measured separately. For example, electric field patterns radiated from microwave horn antennas, electric field intensities coupled through apertures in shielded enclosures, diffraction patterns of electric fields scattered from complicated metallic objects, and electric field modal distributions induced inside cylindrical waveguide cavities were measured. Also, magnetic field distributions near conductive surfaces and induced surface currents on metallic surfaces were determined.

2. IR Measurement Technique (Brief Overview)

This new IR measurement technique is described more fully in the final report of the 1995 AFOSR Summer Faculty Research Program (pages 14-1 through 14-20). A brief summary of the technique is

presented here for completeness.

The IR measurement technique is based on the Joule heating that occurs in a lossy material as an EM wave passes through the material. A thin, planar sheet of lossy carbon loaded paper can be used to map the intensity of the electric field; a thin, planar sheet of lossy ferrite loaded epoxy can be used to map the magnetic field. In either situation, the absorbed heat energy is converted into conducted and convected heat energy and into re-radiated EM energy. The radiated EM energy is concentrated in the long wave (8-12 μm) IR band. This "black body" energy can be detected with an IR (Scanning) Array or with an IR (Starring) Focal Plane Array (FPA).

2.1. IR Experimental Setup

This technique involves placing a lossy non-perturbing IR detection screen in the plane over which the intensity of the EM field is to be measured.

2.1.1. Electromagnetic Parameters

The detector screen is made from a thin sheet of linear, homogenous, and isotropic, but lossy material. From the complex form of Poynting's Theorem for a linear, homogeneous and isotropic material, the absorbed power P_{abs} within a given volume V of the lossy material is a function of the electric (E) and magnetic (H) field intensities inside the screen and is given by

$$P_{abs} = \int_V (\sigma E^2 + \omega \epsilon'' E^2 + \omega \mu'' H^2) dV \quad [W/m^2] \quad (1)$$

where σ is the conductivity of the detector screen, ϵ'' is the imaginary component of the permittivity of the detector screen, μ'' is the imaginary component of the permeability of the detector screen, and ω is the radian frequency of the incident field. The volume integral is over the illuminated portion of the detector screen. The spectral characteristics of the complex constitute parameters (μ, ϵ, σ) of the detector material must be known (or measured) over the entire frequency bandwidth to be measured.

The incident EM energy is absorbed by the lossy material and is converted into thermal heat energy which causes the temperature of the detector material to rise above the ambient temperature of the surrounding background environment by an amount which is proportional to the local electric and/or magnetic field intensity (energy) at each point (pixel) in the screen material. In regions where the field is strong, the absorbed energy is large and the resulting pixel temperatures are high; in regions where the field is weak, the absorbed energy is small and the resulting pixel temperatures are low (near ambient background temperature levels). The resulting two-dimensional temperature distribution over the surface

of the screen can be detected, digitized and stored in the memory of an IR camera. The temperature distribution on the surface of the screen without any EM energy incident on the screen also can be stored in the memory of the IR camera as a ambient background reference temperature distribution. The difference in the temperature at each pixel between the illuminated and the non-illuminated screen is due to the effect of the electric or magnetic field incident on the screen at that pixel location.

The magnitude of the measured EM field can be visualized by presenting the differenced two-dimensional temperature profile as a false color image, where cool colors (for example shades of blue) represent weak areas of EM energy and hot colors (for example shades of red) represent strong areas of EM energy. The resulting two-dimensional false color image is called an IR thermogram, i.e., an iso-temperature contour map, and is a representation of the electric and/or magnetic field distribution in the screen material.

2.1.2. Thermal Parameters

For a planar sheet of detector screen supported by a block of low thermal conducting material, e.g., a styrofoam block, the thermal problem reduces to considering the conductive, convective, and radiative heat losses from the surface of the detector material.

The conductive heat loss h_{cond} is approximated by Fourier's Law of Conductivity

$$h_{cond} = -k \frac{\partial T}{\partial z} \quad [W/m^2] \quad (2)$$

where z is the normal direction into the screen and k is in $W/m \cdot K$.

The convective heat loss h_{conv} is approximated by Newton's Law of Convection

$$h_{conv} = h_o (T - T_{amb})^{1.25} \quad [W/m^2] \quad (3)$$

where h_o in W/m^2K varies between 1.4 and 1.6.

The radiative heat loss h_{rad} is approximated by

$$h_{rad} = \epsilon_{ir} \sigma_{ir} (T^4 - T_{amb}^4) \quad [W/m^2] \quad (4)$$

where ϵ_{ir} is the detector surface emissivity, σ_{ir} is the Stefan-Boltzman constant in $W/m^2 \cdot K^4$, and the temperatures are in degrees Kelvin.

The conductive heat loss is small and can be neglected.

2.1.3. Thermal Equilibrium

The heat transfer problem in the detector material involves solving a non-linear, second order differential equation in both space and time, while considering radiative and convective heat losses from the surface of the material, conductive heat transfer within the material, and the EM power absorbed in the material as a function of distance into the material. For the case of the thin screens considered here, the temperature is initially considered to be constant in the direction normal to the surface of the material, so that the conductive term normal to the surface of the screen can be ignored and the power absorbed can be considered independent of the direction normal to the surface of the screen. Also, the time dependence of the absorbed heat energy is ignored for the steady-state solution.

Relating the convective and radiative heat losses in equations (3) and (4) to the absorbed power in equation (1), results in the following equation at thermal equilibrium:

$$P_{abs} = h_{conv} + h_{rad} \quad (5)$$

For a properly optimized detector screen, thermal equilibrium is achieved in just a few seconds.

This non-linear thermal/electrical equation can be solved for the electric or magnetic field, as a function of the material temperature T , using approximate techniques.

2.2. Approximate Solution

Equation (5) is highly non-linear for large temperature variations above ambient, due to the thermal processes of convection and radiation. However, for small temperature variations of only a few degrees above ambient, equation (5) can be linearized for small incremental temperature changes $\Delta T = T - T_{amb}$ above the ambient temperature T_{amb} .

This condition of small temperature variations above ambient is a desirable operational constraint, since this is also the requirement for small absorption of the EM energy passing through the screen, which equates to a small perturbation of the incident field when performing the measurement. For this minimally perturbing measurement case, an almost direct linear correlation exists between the incremental surface temperature ΔT and the absorbed electric or magnetic field intensity. The incident electric or magnetic field can then be determined from a solution of Maxwell's Equations (Fresnel's and Snell's laws) for an EM wave incident on a planar film of lossy material. Therefore, it is possible to correlate local surface temperature variations ΔT to E or H field intensities.

For the scanning camera available during the extension period at Rome Laboratory to make the IR thermograms, temperature differences ΔT as small as 0.009 °K could be detected.

Care is exercised, therefore, in the selection of the screen material not to significantly perturb the

electric or magnetic field by the presence of the lossy material. The screen is designed to absorb from 1% to 5% of the incident power and to produce a temperature change of less than a few degrees. The constitutive parameters of the IR detector screen can be optimized to produce a large temperature rise in the detector material for a small amount of absorbed energy.

Electric and magnetic fields produced by Continuous Wave (CW) sources operated in the sinusoidal steady-state mode are easy to measure because of the large amount of energy contained in the wave. Transients produced by high power microwave (HPM) pulsed sources, especially repetitively pulsed sources, also can be measured, if the average energy content in the pulse is high enough to raise the temperature of the detector screen material above the minimum temperature sensitivity of the IR camera. The thermal mass of the detector material will hold the absorbed heat energy long enough to capture the IR thermogram of the pulse.

2.3. IR Detector Screen

Referring to equation (1), the screen material can be tailored to respond to only one component of the field, e.g., by optimizing the values of the electrical conductivity σ and the imaginary part of the permittivity ϵ'' of the material relative to the imaginary part of the permeability μ'' of the material, the detector screen can be made sensitive either to the tangential component of the electric field or to the tangential component of the magnetic field in the plane of the screen.

For example, an *electric* field detector screen can be constructed either

- i) from a lossy material with a high conductivity σ and a low imaginary permittivity ϵ'' and a low imaginary permeability μ''
- or ii) from an electrically polarizable material with a high imaginary permittivity ϵ'' and a low conductivity σ and a low imaginary permeability μ'' .

Alternatively, a *magnetic* field detector screen can be constructed from a magnetically polarizable (magnetizable) material with a high imaginary permeability μ'' and a low conductivity σ and a low imaginary permittivity ϵ'' .

The optimization of the thermal and electrical parameters of the detection screen material should be guided by a thermal/electrical computer code based on a plane wave normally incident on a planar interface between air and the lossy detector material. Other absorptive and re-emittive transducing materials have been studied for use as passive thermal screens for IR thermograms.

2.3.1 Electric Field Detector Screen

For the detection of electric fields, the IR detector screen can be made from a planar sheet of lossy thin-film material. Several different detector screens were used to make electric field thermograms. One was a carbon loaded thin film (e.g., Teledeltos Paper) 80 μm thick with a conductivity of 8 mhos per meter. The other screens were made from carbon-loaded Kapton films. The films were loaded with different resistivities per square. These materials are non-polarizable and non-magnetizable; therefore, the imaginary components of the permittivity ϵ'' and the permeability μ'' are negligibly small. For these conducting, non-polarizable, non-magnetic screen materials, maximum heating occurs due to the electric field and negligible heating occurs due to the magnetic field.

For plane waves normally incident on this carbon loaded electric field detection screen and for an IR FPA with a temperature sensitivity of approximately 0.01 $^{\circ}\text{K}$, electric fields with a magnitude on the order of 61.4 V/m (1 mW/cm² of incident power) can be detected. This result was obtained empirically from experimental data in which the incident power level was incrementally decreased until no electric field contour lines on the IR thermogram were discernable from the ambient background level.

The carbon loaded Kapton and Teledeltos papers were compared for the thermal properties. A wide range of absorption coefficients are available in each material.

2.3.2 Magnetic Field Detector Screen

For the detection of magnetic fields, an IR detection screen can be fabricated using a ferrite loaded epoxy. Several different ferrite powders were tested, viz. Nickel Zinc Copper Ferrite, Iron Silicide Ferrite, Ferrite 50, etc. The best detector screen used to make magnetic field thermograms had an 80% by weight mixture of Ferrite 50 with a thickness of 0.5 mm and an imaginary relative permeability μ_r'' of approximately 2. The imaginary component of the permittivity ϵ'' and the conductivity σ were negligibly small. For this magnetic, non-conducting, non-electric screen material, maximum heating occurs due to the magnetic field and negligible heating occurs due to the electric field.

In the fabrication of a magnetic screen, it is difficult to keep the electric polarizability of the epoxy base material from contributing to the imaginary component of the permittivity ϵ'' of the composite and, thereby, introducing additional electric polarization losses. If the electric properties of the ferrite are not carefully controlled and minimized, the absorbed power will be due to both the electric and magnetic components of the EM field, as given by equation (1). In this case, it is difficult, if not impossible, to separate the two coupling mechanisms from each other and to calibrate the screen.

To detect surface currents on a metallic structure, the magnetic detection screen can be placed in direct contact with the surface of the metal. The thickness of the detector screen must be kept to a

minimum, so that the magnetic field that is measured on the outer surface of the screen is a direct reflection of the magnetic field on the inner surface of the screen in contact with the surface of the metal. The material can be held flat against the surface of the metal using a non-conducting glue.

The surface current is related to the tangential component of the magnetic field intensity on the surface of the metal by the magnetic boundary condition on a perfect conductor:

$$\hat{n} \times H|_s = J_s$$

where \hat{n} is the outward normal to the surface S . The magnetic field is perpendicular to the direction of the surface current.

A new magnetic material composed of ferro-electrical composites also was tested. The initial tests showed that the imaginary permeability μ'' of the material was too low to strongly couple the magnetic field to the detector material to produce a good IR thermogram. A thicker material with a higher imaginary μ'' will be built for future tests.

2.4 IR Camera

The temperature difference between the screen material and the background is detected, digitized, and stored in the memory of an IR camera on a pixel by pixel basis. The IR imaging system used to take the IR thermograms has approximately 200 by 100 pixels per frame of data. The detector is scanned over the image. The detector is a Mercury-Cadmium-Telluride (HgCdTe) IR photo detector operated in a photo voltaic mode. The detector operates at Liquid Nitrogen temperatures. The camera can detect temperature differences of approximately 0.009 °K, and has a relative accuracy of plus or minus 10% when detecting EM fields.

2.5 IR Images

The stored IR thermogram represents the temperature distribution over the extent of the detector screen and is a map of the intensity of the electric or magnetic field distribution absorbed in the screen. For small temperature rises less than a few degrees above ambient, the electric and magnetic field intensities are nearly linearly proportional to the temperature change.

2.5.1 Spatial Resolution

The spatial resolution of the IR thermogram is a function of the number of pixel elements in the IR

camera, and is fixed by the angular resolution of the telephoto, regular, or wide angle lens used on the IR camera when making the IR image. A telephoto lens can be used to look at small details in the field structure on the detector screen; a wide angle lens can be used to look at large scale trends in the field structure in the detector screen.

The telephoto lens also has the added advantage for regular field mapping applications of allowing the IR camera to be located far away from the object under test and, thus, removing any perturbing effects that the metallic structure of the camera might have on the field distribution being measured.

2.5.2 Thermal Resolution

The thermal resolution of the IR thermogram is a function of the digitizer in the IR camera. The IR camera used to take the IR thermograms has a 12 bit digitizer. For a 12 bit digitizer, the temperature range seen by the IR camera is divided into 256 increments. Each digitized increment is assign a unique color, resulting in a temperature resolution of 256 color levels.

2.5.3 Thermogram Errors

The resulting IR image of an EM field depends on the combined EM and thermal properties of the detector material and is subject to several significant, but controllable, errors.

2.5.3.1 Lateral Conduction Effects

Conductive heating in the transverse direction within the screen material causes thermal "bleeding" from the hot spots on the screen to nearby cold spots. This thermal bleeding tends to fill in the minimums (nulls) somewhat, whereas, the areas of maximum heating (peaks) are not affected very much by the bleeding effect. This effect can be minimized by operating at small temperature variations above ambient.

2.5.3.2 Lateral Convection Effects

Convective heating of the top of the screen from heat rising from the bottom of the screen causes the top of the screen to appear slightly hotter than the bottom of the screen. This "blurring" of the image can be kept to a minimum by operating at small temperature variations above ambient.

This blurring effect can be eliminated completely by placing the IR detector screen in a horizontal position and observing the image with the IR camera looking down on the screen from above or from the

side using an IR mirror.

2.5.4 IR Measurement Accuracy

The accuracy of the IR measurement technique was demonstrated by performing two simple experiments with known theoretical solutions. In one experiment, the diffraction pattern from a "Lloyd's Mirror" was measured. In the other experiment, the diffraction pattern from an "knife edge" conducting half-plane was measured.

In the Lloyd's mirror experiment, the resulting diffraction pattern is due to the antenna interfering with its image in a large ground plane. The near field (Fresnel Zone) antenna pattern of the horn antenna was used to obtain the theoretical results.

In both experiments, the screen material was optimized to measure only the tangential component of the electric field intensity in the plane of the screen.

These experiments were performed in an anechoic chamber. Good correlations between theory and experiment were obtained. The worst errors occurred in the minimums (deep nulls) of the diffraction patterns were thermal bleeding from the surrounding hot areas tended to obscure the real depth of the minimums. Some thermal bleeding out of the maximums into the surrounding areas also occurred, obscuring the real height of the maximums.

Even with conductive bleeding and convective blurring of the image, the measurement error was less than approximately 10% under controlled test conditions.

2.6 IR Advantages and Disadvantages

The IR measurement technique provides a quick and accurate method to observe EM fields in a two-dimensional plane. However, only the magnitude of the electric or magnetic field is measured; no phase information is detected. Also, since this technique is based on the thermal mass of the detector material, high energy is required to produce good thermal images of EM fields.

3. IR Thermograms

In previous applications of the IR measurement technique to EM problems, only the relative magnitude of the EM fields were determined and no phase information was obtained. In this extension period, absolute magnitudes and relative phase information was determined.

3.1 Magnitude Estimates (Calibration)

The IR measurement technique was calibrated during the extension period at Rome Laboratory.

3.1.1 Relative measurements

The relative accuracies of the field intensities measured by the IR thermograms taken with the IR scanner were determined using numerical computer codes to predict the normalized field intensities for the same geometries. The relative accuracies were never greater than 10% in error. These errors occurred at the minimum (null) field intensity positions in the thermograms, where the effect of thermal conduction (bleeding) from adjacent hot spots tends to fill in the nulls somewhat. Overall, at positions away from nulls, the technique was closer to only 1% in error. Therefore, the technique has the potential to produce extremely accurate field intensity measurements, much less than ± 1 dB in error.

3.1.2 Absolute Measurements

To determine the absolute electric or magnetic field strengths from IR thermograms, the IR scanning detectors and detector screens were calibrated.

Appropriate theoretical models and experimental procedures are being developed to permit the absolute determination of electric and magnetic field intensities from thermographic images (IR thermograms) of microwave fields.

In the extension period, an initial attempt was undertaken to calibrate the IR thermogram images to determine the absolute magnitude of the electric or magnetic field intensity. Several empirical tests were performed with a standard gain horn aperture antenna to correlate the intensity of the EM power of the radiated wave incident on the detector screen with the intensity of the measured temperature (color) of the screen on a pixel by pixel basis. The measurement procedure developed here will be repeated, after the test is automated to improve the accuracy of the measurement procedure and the repeatability of the data.

A purely theoretical effort based only on the thermal and electrical parameters of the detector screen also will be undertaken to predict the incident electric field intensity as a function of the measured temperature.

3.1.2.1 Theoretical Calibration

A thermal/electromagnetic model is being developed for the interaction of microwaves with a lossy/complex absorber material. The theory is based on the complex form of Poynting's Theorem for an

absorbing material with complex constitutive parameters. The broad-band frequency dependence of the complex constitutive parameters of the IR detection screen material must be determined and used in the model of the detector screen along with its thermal properties.

3.1.2.2 Experimental Calibration

The intensity levels (equi-color levels) of the IR thermograms were empirically calibrated using different standard gain horn antennas at several frequencies, angles of incidence, and polarizations in the near and far fields of the antenna. The predicted field level is simply associated with the resulting color level.

3.1.2.2.1 Experimental Setup

As the first step in calibrating the detector screen, an experiment was performed on a planar sheet of the detector material. In the experiment, the screen was positioned in front of a standard gain horn antenna and oriented in several different planes making different angles relative to the antenna. For one series of IR thermograms the screen was positioned in the axial plane in the near and far fields of the horn. For this case, the normal to the screen was oriented perpendicular to the direction of the incident radiation (perpendicular incidence). The camera was placed in front of the screen looking at the thermal pattern on the screen. In other tests, the normal to the screen was rotated to different angles relative to the bore-sight of the horn. The frequency, angle of incidence, and polarization of the incident wave were varied.

3.1.2.2.2 Experimental Data

The induced temperature distribution in the plane of the detector screen was mapped with the IR camera. The measured temperature on the bore-sight of the horn antenna was compared to the value of the incident field calculated using the Friis Transmitter/Receiver Formula with the known and/or measured characteristics of the antenna, corrected for all losses and mismatches. Forward and reverse power were measured with calibrated power meters. The distance to the screen from the phase center of the horn and the angle of the screen relative to the bore-sight of the horn were measured for each configuration.

A MathCad computer program was written to predict the incident field in the far field of the antenna. Due to power constraints, which are common occurrences with the IR technique in order to produce good IR thermograms, some tests were actually performed in the near field of the antenna. A near-field correction routine was then implemented to correct for the near-field effects.

Experimentally obtained temperatures from the IR thermograms were plotted against the incident

power densities for each case. A series of thermograms were taken in the 1 to 2 GHz range of frequencies at distances from 1 to 2 meters from the antenna. The background ambient temperatures for each case were also recorded. Many other configurations were tested at different distances and at various other frequencies, polarizations, and angles of incidence.

On the average, the corrected temperature differences between the various cases was less than ± 0.5 °F. This temperature error translates into a ± 1 dB field measurement error.

3.2 Phase Estimates

The IR measurement technique provides a quick and accurate method to observe EM fields in a two-dimensional plane. However, only the magnitude of the electric or magnetic field can be measured directly; no phase information is detected. In the extension period, several plans were devised to implement a phase measurement capability.

Several different techniques were considered. One technique is based on energy minimization and the other technique is based on microwave holography.

3.2.1 Energy Minimization Technique

A numerical global-energy minimization technique (based on the Method of Moments) can be used to estimate the desired phase information. This part of the work is in conjunction with Professor Tapan Sakar at Syracuse University. Using the IR measurement technique, which is especially well suited for near-field measurements where the energy density is high, near-field to far-field transformations can be performed once the estimated phase information is included with the standard magnitude data.

3.2.2 Microwave Holography Technique

A microwave holographic technique also can be used to estimate the desired phase. This part of the work is in conjunction with Dr. Carl Stubenrauch at the National Institute of Standards & Technology (NIST/Boulder).

Two distinctly different techniques within the broad area of holography will be attempted. One is based on a iterative Plane-to-Plane technique, the other is based on an IR hologram.

3.2.2.1 Plane-to-Plane (PTP) Phase Retrieval Technique

One technique is a iterative Plane-to-Plane (PTP) Fourier transformation. In this technique, the

measured magnitude is used to correct an iterative procedure that estimates the phase of the radiation on two different planes in the near field of the antenna. To increase the convergence rate of the iterations, the phase data is transformed between the two planes with intermediate transformations to the aperture plane of the antenna, where the magnitude and phase data are truncated outside the immediate aperture plane of the antenna.

3.2.2.2 Microwave Holographic Phase Retrieval Technique

The other technique is to make an IR hologram and to read out the images by simulating the data on the computer. This technique can be further enhanced by making several holograms which can be simultaneously readout on the computer to improve the accuracy of the results. Two different holographic planes can be used, or four different holograms made in one plane, but each taken with four different phase shifts introduced between the antenna under test and the reference readout antenna.

The accuracy and easy of implementation of these different techniques will be compared in future work.

4. Results

The electric or magnetic field distributions can be predicted for a well-known, theoretically tractable problem geometry, e.g. scattering from a linear wire antenna, scattering from a thin rectangular slot aperture, antenna radiation in the near or far field of a rectangular horn antenna, induced cavity modes in a rectangular or cylindrical waveguide, etc. The scattering predictions can be made for a normally or obliquely incident plane wave as a function of frequency, angle of incidence, and horizontal/vertical polarizations. These theoretical predictions can be verified experimentally to validate the IR/EM interaction model for each problem geometry.

Examples of IR measurements taken during the extension period at Rome Laboratory are now presented.

4.1 Scattering from a Cylinder

Experimental tests were performed on a conducting cylinder irradiated by an incident plane wave. E- and H- field patterns of the scattered/diffracted energy from the cylinder were measured. Tests were conducted at various microwave frequencies relative to the resonant frequencies associated with the cylinder and at numerous angles of incident (e.g., end-on, broad-side, oblique-incidence) and for several

different polarizations of the incident field relative to the axis of the cylinder (e.g., horizontal, vertical, and skewed). Each test was performed in the near and far fields of the antenna. IR thermograms of the scattered fields were taken. The diffraction patterns of the electric field scattered from the cylinder are clearly indicated in these thermograms.

The equi-temperature contour levels in the IR thermograms are being compared to numerical predictions of the scattered electric field intensities. The numerical predictions are being performed by Dr. Zuffada at Caltech's Jet Propulsion Laboratory (JPL). These results will be used to verify JPL's numerical code which combines the Method of Moments (MoM) with a Finite Element Method (FEM) to determine electric field intensity levels.

4.2 Scattering from an F16 Scale Model Aircraft

The IR imaging technique was used to map the fields around a simple generic model of an F16 aircraft. A plastic F16 scale model was constructed and sprayed with several coats of silver paint to make it conductive. IR thermograms of the magnitude of the scattered electric field intensity were taken in the horizontal and vertical longitudinal planes through the fuselage of the aircraft. The diffraction patterns of the electric field scattered from the aircraft and the scattering centers are clearly indicated in these thermograms. Standing waves are setup between the incident wave and the reflected wave off the aircraft. Surface waves are established along both sides of the fuselage and the front edges of the wings between the nose of the aircraft and the missile rails. Diffraction occurs off the wing tips. A shadow zone appears behind the aircraft.

The equi-temperature contour levels in the IR thermograms are being compared to numerical predictions of the scattered electric field intensity. The numerical predictions are being performed with the GEMACS code at Rome Laboratory. These results are being used to validate the accuracy of the GEMACS numerical code (developed to predict scattered and diffracted fields from advanced aircraft configurations).

4.3 Phased Array Antenna Patterns and Element Efficiency

It is difficult and time consuming to determine which elements of a large phased array antenna are not functioning at full capacity or which elements are not working properly using standard B-Dot or D-Dot probes. Due to the 2D mapping capabilities of the IR technique, this technique was used to map the close-in near-field just above the aperture plane of a large array and to use this near-field image to determine the state of the array for different operational modes. The near-field image revealed, directly on observation, whether or not an element of the array was working. This simple technique could lead to a

great cost and time savings in the testing of large Air Force phased array radar systems and will be developed in future work.

5. Summary of Extension Period Tasks

1. Test were performed on several IR detector screen materials to determine which ones produce a large temperature rise in the detector material for a small amount of absorbed energy.
2. A magnetic field detection screen, which used either a ferrite loaded epoxy film or a new "spin-on" material technique, was developed.
3. A phase measurement capability is being developed using a numerical global-energy minimization technique (based on the Method of Moments) or one of several holographic techniques to estimate the desired phase information
4. IR thermograms were calibrated to determine the absolute magnitude of the electric field intensity using empirical experimental data.
5. IR thermograms of the scattered microwave energy for an advanced model aircraft are being used to verify the accuracy of the GEMACS numerical code.
6. IR thermograms of the scattered microwave energy from a cylinder are being used to verify the accuracy of the JPL FEM/MoM numerical code.
7. IR thermograms of the close-in near-field just above the aperture plane of a large phased array were used to determine the operational state of the elements of the array for different operational modes.

6. Conclusions

Initial empirical results from anechoic chamber tests indicate that the induced temperature distribution across the IR thermogram in a properly designed, lossy, planar IR detection screen placed in the area over which an EM field is to be mapped can be calibrated to measure the incident EM wave using the Joule heating of the screen material.

The IR measurement technique is, therefore, a viable method to aid in the determination of EM fields in various test situations. The IR method allows for rapid observation of EM field activity and interference, resulting in an in-depth understanding of EM scattering phenomena. An experimental technique is being used to measure and calibrate the absolute magnitude of the field intensity. A theoretical approach to this calibration problem is also being undertaken.

The IR measurement technique is a viable method to aid in the determination of EM fields scattered

from complex metallic objects and the EM energy coupled into complex cavity structures. This method is of particular importance in the study of scattering surfaces with complicated geometrical shapes, whose patterns of surface current distributions may not be found easily using theoretical methods.

The IR method allows for rapid observation of EM field activity and interference, resulting in an in-depth understanding of the EM scattering and coupling phenomena. Qualitative and quantitative comparisons can be made between the fields measured using the thermal radiation experimental approach and the fields predicted using a theoretical/numerical approach. Experimental and theoretical data, therefore, can be easily correlated with this technique.

7. Future Work

The development of an optimized magnetic detection screen is in progress. The use of spin-on magnetic thin films is being investigated.

It is also possible to estimate the phase of the incident wave by making a numerical model of the thermal/electromagnetic interaction in the detector screen material and to use energy minimization techniques to estimate the phase of the incident wave. Microwave holographic techniques also can be used to determine the phase of the wave incident on the IR detector screen if the magnitude of the incident electric field is measured at several different positions in the near field of the radiation source. This phaseless measurement work is being done in conjunction with Syracuse University (Prof. Tapan Sakar) and the National Institute of Standards and Technology (NIST) (Dr. Carl Stubenrauch).

8. Publications

The papers published on this project during the AFOSR Summer Research Extension Program are listed in the Appendix.

Acknowledgments

The author would like to acknowledge the support of the AFOSR Summer Research Extension Program on this IR project and the help of Michael Seifert at Rome Laboratory. Also, due to extra funding from AFOSR through Seiler Labs at the US Air Force Academy, several trips were taken to Rome Labs to perform the IR experiments. The author also gratefully acknowledges that extra support.

APPENDIX

A. *Papers Presented:*

1. J.D. Norgard
"Measurement of Absolute Electromagnetic Field Magnitudes using Infrared Thermograms"
Proceedings of the Quantitative Infrared Measurement Conference
January 1995.
2. J.D. Norgard, R.M. Sega, M.F. Seifert, and T. Pesta
"Measurement of Absolute Electromagnetic Field Levels using Infrared Thermograms"
Proceedings of the URSI Winter Meeting
(U of Colorado), Boulder, CO
January 1995.
3. J.D. Norgard, R.M. Sega, M.F. Seifert, A. Pesta, and T. Blocher
"Code Validation of Aircraft Scattering Parameters Using IR Thermograms"
Proceedings of the ACES Conference
NPS, Monterey, CA
March 1995.
4. J.D. Norgard, R.M. Sega, M.F. Seifert, and A. Pesta
"Empirical Calibration of Infrared Images of Electromagnetic Fields"
Proceedings of the SPIE/Thermosense XVII International Conference
Orlando, FL
May 1995.
5. J.D. Norgard, M.F. Seifert, and T. Pesta
"Infrared Images of Electromagnetic Fields ... Relative and Absolute Calibration"
Proceedings of the Dual-Use Technology Conference
SUNY/Utica, NY
May 1995.

6. J.D. Norgard and M.F. Seifert
"Infrared Images of Electromagnetic Interference (EMI) Coupled through a Thin Slot Aperture in a Cylindrical Waveguide Cavity / Part I: Theory"
Proceedings of the IEEE/EMC Conference
Atlanta, GA
August 1995.
7. J.D. Norgard and M.F. Seifert
"Infrared Images of Electromagnetic Interference (EMI) Coupled through a Thin Slot Aperture in a Cylindrical Waveguide Cavity / Part II: Experiment"
Proceedings of the IEEE/EMC Conference
Atlanta, GA
August 1995.
8. J.D. Norgard, R.M. Sega, and M.F. Seifert
"Electromagnetic Field Measurements using Infrared Imaging Techniques"
Proceedings of the ICEAA Conference
Polytecnico Di Torino
Turin, Italy
September 1995.

B. Seminars Presented:

1. "Calibration of Infrared Images of Radiated Electromagnetic Fields"
US Air Force Academy
Department of Electrical Engineering
Colorado Springs, CO
June 1995
2. "Infrared Images of Electromagnetic Fields"
Norwegian Defense Research Establishment (NDRE)
FFI/Geophysics
Kjeller, Norway
June 1995

C. *Journal Articles Published:*

1. "Measurement of Absolute Electromagnetic Field Magnitudes using Infrared Thermograms"

EuroTherm Quantitative Infrared Measurement Journal

January 1995

D. *Papers Submitted:*

1. "Measurement of the Relative Phase of Electromagnetic Fields using Infrared Thermograms"

URSI Winter Meeting

CU/Boulder, CO

January 1996

2. "Code Validation of Cylindrical Scattering Parameters using IR Thermograms"

ACES Symposium

US NPS/Monterey, CA

March 1996

3. "Phase Measurements of Electromagnetic Fields using Infrared Imaging and Microwave Holography"

SPIE Conference

Thermosense XVIII

Orlando, FL

April 1996

4. "Near Field Phase Reconstruction using Plane-to-Plane Iterative Fourier Processing and Infrared Thermograms of Electromagnetic Fields"

NEM Conference

Albuquerque, NM

May 1996

5. "Near-Field to Far-Field Antenna Pattern Measurements using Infrared Imaging and Microwave Holography Techniques"

Dual Use Conference

Syracuse University
Syracuse, NY
May 1996

6. "Microwave Holography using Infrared Images of Electromagnetic Fields"

AMTA
Montreal
June 1996

7. "Complex Electromagnetic Magnitude and Phase Measurements using Infrared Imaging and Microwave Holography"

PIERS Conference
Innsbruck, Austria
July 1996

E. Trips to Rome Laboratory:

Due to extra funding from AFOSR through Seiler Labs at the US Air Force Academy, a number of trips (approximately one a month) were taken to Rome Labs to perform some of the IR experiments.

Femtosecond Pump-Probe Spectroscopy System for the Design and Characterization
of High-Speed All-Optical Modulators at $1.3\ \mu\text{M}$

Dean Richardson
Assistant Professor
Department of Electrical Engineering Technology

SUNY Institute of Technology
Utica, NY

Final Report for:
Summer Research Extension Program

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base
Washington, DC

and

Rome Laboratory

December 1995

FEMTOSECOND PUMP-PROBE SPECTROSCOPY SYSTEM FOR THE DESIGN AND CHARACTERIZATION
OF HIGH-SPEED ALL-OPTICAL MODULATORS AT 1.3 μM

Dean Richardson
Assistant Professor
Department of Electrical Engineering Technology
SUNY Institute of Technology @ Utica/Rome

Abstract

A titanium-doped sapphire (Ti:S) regenerative amplifier system was constructed to amplify femtosecond optical pulses from a Kerr-lens mode-locked Ti:S oscillator utilizing the chirped-pulse amplification technique. The ~ 100 femtosecond, 5 nanojoule pulses were first stretched temporally by a factor of approximately 1000, then coupled into the amplifier using a Faraday isolator, a Brewster-cut polarizer, and a Pockels E-O modulator. After undergoing approximately 15 round trips within the amplifier, the pulses were extracted at a rate of 1 kHz, emerging with an energy of roughly 0.5 millijoules. The amplified pulses were suitable for recompression and subsequent pumping of crystalline quartz for the purpose of femtosecond continuum generation, which is a crucial prerequisite to performing femtosecond pump-probe spectroscopy in semiconductor-based modulator devices.

FEMTOSECOND PUMP-PROBE SPECTROSCOPY SYSTEM FOR THE DESIGN AND CHARACTERIZATION OF HIGH-SPEED ALL-OPTICAL MODULATORS AT 1.3 μM

Dean Richardson

Introduction

The design and optimization of Fabry-Perot etalon-based high-speed semiconductor modulators requires an accurate understanding of the nonlinear absorption and refractive-index dynamics of the modulator cavity on ultrafast timescales. Femtosecond pump-probe spectroscopy is viewed by many researchers as the preferred method for characterizing the temporal behavior of these carrier-density dependent optical nonlinearities in semiconductors¹. In this approach, the differential transmission response of a potential modulator material to a short, intense, narrowband pump pulse is sampled over a range of delay times with femtosecond resolution using a weaker, broadband femtosecond continuum pulse. Using this technique, the entire band-edge absorption spectrum (and through the Kramers-Kronig transformation, the corresponding index changes) can be tracked simultaneously as a function of time after the onset of short-pulse excitation. These parameters can then be utilized in developing the optimum cavity structure for high-contrast, low insertion-loss modulators.

Generation of the broadband, "white light" femtosecond continua utilized in pump-probe spectroscopy typically requires extremely high-energy pulses to be focused onto an ethylene glycol jet or onto a quartz or glass substrate. Early efforts² in this regard relied on colliding-pulse-mode ring cavities and multiple dye-jet amplifier stages pumped by bulky, temperamental copper-vapor lasers; fortunately, compact, all-solid-state solutions have recently become practical³. The most common of these makes use of a Ti:Sapphire regenerative amplifier, usually pumped by a Q-switched, frequency-doubled Nd:YAG/YLF laser, to amplify femtosecond-scale pulses from a Ti:Sapphire oscillator to energy levels sufficient for continuum generation.

Though such systems are now commercially available, they remain prohibitively expensive for all but the most well-funded research laboratory. Thus our project focused primarily on the design and construction of such a system from scratch, with the goal of generating the highly-amplified pulses needed for femtosecond continuum generation.

Discussion of Problem

Our home-built regen system eventually consisted of the following key sub-systems, diagramed in Fig. 1:

- Coherent 310 ten-watt Argon-Ion pump laser
- Ti:Sapphire oscillator, with intracavity dispersion-compensating prisms
- Pulse-stretching assembly, including a 1:1 telescope and a high-efficiency precision-ruled grating
- Ti:Sapphire regenerative amplifier cavity, with intracavity Medox Pockels Cell and high-performance calcite polarizer for insertion and extraction of seed and amplified pulses
- Q-switched, frequency-doubled Nd:YLF pump laser, operating at 1 kHz.

Discussion of Problem (cont.)

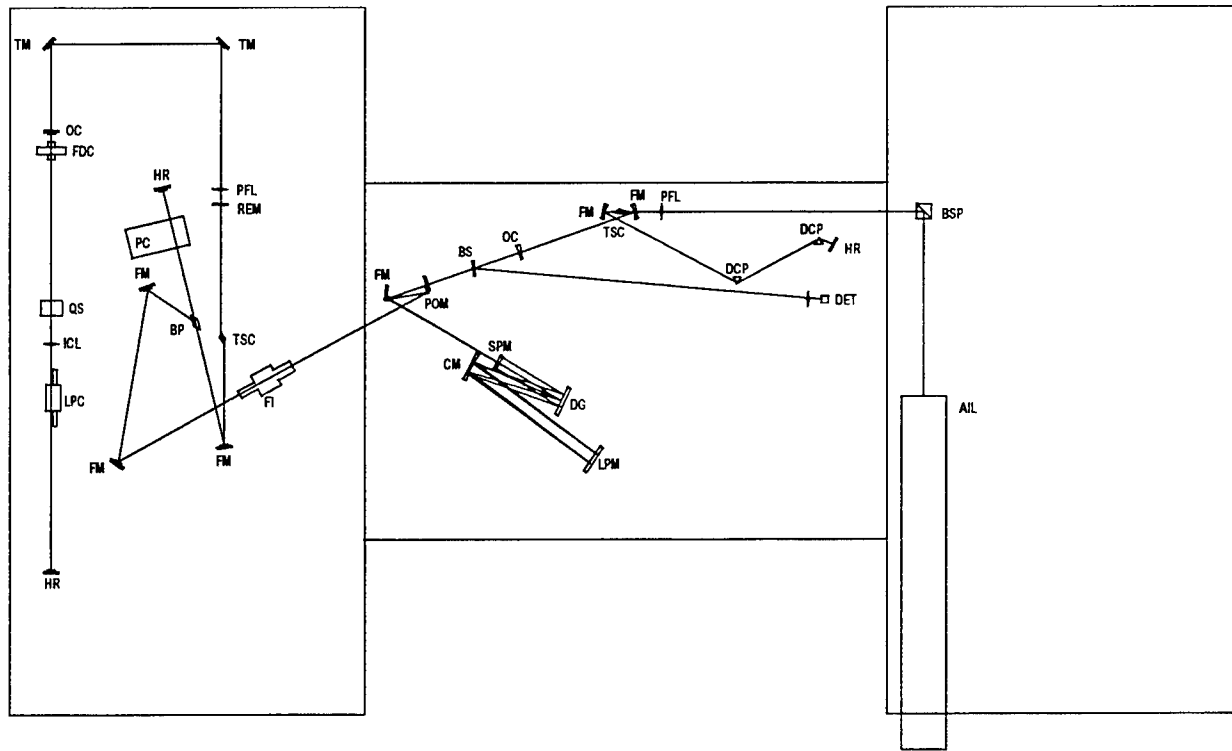


Figure 1 Continuum generation system layout. AIL -- argon-ion laser; BSP -- beam-steering periscope; PFL -- pump focus lens; FM -- fold mirror; TSC -- Ti:Sapphire crystal; DCP -- dispersion-compensating prism; HR -- high-reflector; OC -- output coupler; BS -- beamsplitter; DET -- detector; DG -- diffraction grating; CM -- curved mirror; LPM -- large plane mirror; SPM -- small plane mirror; POM -- pick-off mirror; FI -- Faraday isolator; BP -- Brewster polarizer; PC -- Pockels cell; REM -- regen entrance mirror; LPC -- laser pump chamber; ICL -- intracavity lens; QS -- Q-switch; FDC -- frequency-doubling crystal; TM -- turning mirror.

The basic theory of operation for each of the system components can be outlined as follows:

- The Argon-Ion laser acts as a pumping source for a Ti:Sapphire crystal, creating a population inversion and accompanying gain within the oscillator cavity.
- Through the phenomenon of Kerr-Lens-Modelocking (KLM) or “thermal lensing”, the Ti:Sapphire crystal is induced to generate a periodic buildup and release of optical energy. In concert with the resonator geometry, this mechanism leads to the pulsed mode of operation being favored over CW operation.
- A suitably-aligned prism-pair generates a controllable amount of negative group-velocity dispersion, allowing for the canceling of the dispersion induced by the Ti:Sapphire crystal and mirrors and thus for the generation of femtosecond-scale pulses by the oscillator.
- Prior to amplification, the pulses are stretched in time so as to reduce their peak intensity, thus preventing damage to key components of the amplifier system. This is done in such a fashion as to preserve the various spectral components of the pulse with near-perfect fidelity so as to allow for later re-compression back to femtosecond timescales.

Discussion of Problem (cont.)

- The stretched pulses are coupled into the regenerative amplifier cavity via a Faraday isolator and an intracavity Brewster-cut calcite polarizer. The isolator rotates the polarization of the input pulses from horizontal to vertical, so they can be reflected from the Brewster polarizer into the regen cavity.
- The Ti:Sapphire crystal within the regen cavity is pulse-pumped by a Q-switched, frequency-doubled Nd:YLF laser. The energy from the pump pulses creates a population inversion in the crystal, allowing stretched pulses admitted from the oscillator to be amplified each time they pass through during the timespan of the pump pulses.
- The regen cavity also contains a Pockels cell, which can rotate the pulses' polarization under electrical control on a nanosecond timescale. During the time intervals that the Nd:YLF pump laser turns the regenerative amplifier "on", two different voltages are applied to the Pockels cell, first allowing the admitted pulses to make multiple passes through the amplifier, then allowing a particular amplified pulse to exit the cavity via reflection from the Brewster polarizer.
- The Faraday isolator then rotates the polarization of the exiting amplified pulse and redirects it, insuring that this pulse does not return to the stretcher or oscillator, but instead becomes available to the compressor assembly as desired.

The bulk of our effort was spent aligning and optimizing each of these sub-systems separately, then getting them to function together. As was to be expected, the operation of system components positioned towards the end of the beamtrain (i.e., the regen cavity itself) depended crucially on the performance of initial system elements such as the Argon-Ion laser and the Ti:Sapphire oscillator. So though satisfactory mode-locked operation of the oscillator was achieved early in the project, we had to re-align and re-optimize it each time new elements were added to the system. Further, both the oscillator and the Argon-Ion laser experienced severe degradation in output quality over the course of the project, necessitating significant repairs and component substitutions on several occasions. Nonetheless, the system eventually functioned as intended, producing ultrafast amplified pulses at the millijoule levels required for continuum generation.

Methodology

In this section we describe in detail our system-building efforts, component by component, including obstacles overcome and lessons learned.

Argon-Ion Pump Laser

This laser had been purchased previously but had seen little use prior to this project. The Coherent Innova 310 is touted as providing rock-solid pumping capability for ultrafast systems. It is an all-lines laser, producing a peak CW output of 10 W in the 488-514 nm range. It includes an active-stabilization mechanism (trade name: "PowerTrack") which according to product literature facilitates long-term power and beam-pointing stability by dithering the rear cavity mirror when needed using an internal feedback loop. Despite this capability, we found significant long-term drift in the laser power over the

Methodology (cont.)

course of a day's operation to be a frequent problem, and we eventually turned the PowerTrack system off for good.

More crucial, however was the gradual degradation in the quality of the spatial mode structure of the output beam which we observed about midway through the project. Argon laser tubes are furnished with a quartz window aligned at Brewster's angle at each tube end. We found that the output beam of our laser had gradually generated a burn spot in this window on axis, degrading the TEM_{00} output and favoring the TEM_{10} and TEM_{01} modes, which have no on-axis amplitude. As a result, we observed the donut-shaped output mode which is formed by the sum of these two higher-order modes. This made pumping of the Ti:Sapphire crystal unreliable and its output unstable. In particular, we could not maintain self-mode-locked operation for more than a minute or two at a time with the Argon-Ion laser behaving this way. The oscillator would frequently experience the so-called "CW breakthrough" phenomenon, in which the train of pulses with a broad spectral bandwidth would shut off and be replaced with a narrow bandwidth CW spike.

To rectify this situation, we sought Coherent's help, and learned of a procedure whereby we could apply an etchant known as "ABF" to the quartz window to eat away the pitted surface, leaving the window smooth and clean as before. Though this seemed to improve the output quality for a time, the "donut mode" behavior soon appeared again, and we requested a service call. Several weeks passed before a visit could be arranged, but when it was completed, reasonably good-quality output from the Argon-Ion laser was finally restored. These problems cost our system several months of downtime while the various problems were diagnosed, and slowed the pace of alignment efforts for both the stretcher and the regen significantly.

Mode-locked Ti:Sapphire Oscillator

Our Ti:Sapphire resonator design was modeled after descriptions by Murnane et al. at Washington State University of their pioneering x-cavity configurations and component choices⁴. The mirror set, purchased from CVI, consisted of two fold mirrors, each with a 100 mm radius of curvature, a high-reflector, and a 5% output coupler, all coated for a center wavelength of 780 nm. The fold mirrors were coated to admit the pump laser light at 488/514 nm, while the output coupler was slightly wedged for better performance. The fold mirrors were tilted so as to produce a fold angle for the beam of approximately 18°.

For the gain medium we used a 1 cm long, 4 mm diameter cylindrical Brewster-cut Ti:Sapphire crystal, purchased from Union Carbide crystal products division. The initial crystal had an absorption (α) value of 2.5 per cm, with a figure-of-merit > 200. The crystal was enclosed in a custom-machined holder consisting of a two-piece cylindrical copper jacket through which two paths were provided for cooling water, held in an aluminum retaining bracket. The Argon-Ion pump light was focused onto the Ti:S crystal using a 91 mm focal length plano-convex lens.

Methodology (cont.)

For dispersion compensation, we used a pair of SF10 isosceles Brewster prisms, separated by approximately 28 cm. The prism pair was located in the long arm of the x-cavity, near the high-reflector, as shown in Fig. 1.

For diagnostic measurement of the mode-locked pulse train, and for later synchronization of the intracavity Pockels cell triggering for the regen, a small portion of the oscillator output signal was picked off using a 5% beamsplitter and redirected to a 200-ps risetime reverse-biased photodiode detector. During mode-locked operation of the oscillator, a train of pulses was readily visible on our 250 MHz oscilloscope, as shown in Fig. 2. The repetition rate is easily calculated from the oscilloscope trace to be 100 MHz, indicating that the cavity was quite close to 1.5 meters in length.

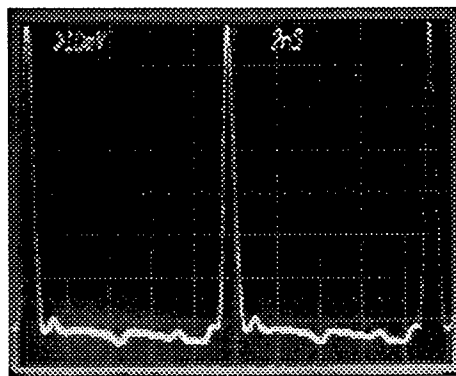


Figure 2 Oscilloscope trace showing signal from 200-ps risetime detector for mode-locked oscillator output. Repetition rate is roughly 100 MHz.

Pulse-Stretching Assembly

Though the goal of our regenerative amplifier system was to produce high peak-power femtosecond pulses, it was obviously not practical to perform the amplification directly on the femtosecond pulses from the oscillator in a single step. The intensity of the amplified pulses would very likely exceed the damage threshold for the mirrors, coatings, and crystal used in the amplification process. Instead, the pulses were first “stretched” by a factor of roughly 1000 from a FWHM of ~ 100 fs to a FWHM of ~ 100 ps through the process of chirped-pulse amplification⁵ (CPA), keeping their energy constant but maintaining the peak intensity per pulse at safe levels. In the CPA approach, the chirp must be carefully controlled and monitored using autocorrelation or other diagnostics, so that the pulses can be compressed back to their original femtosecond timescale once the amplification process is completed. This we were successful in doing, as subsequent results indicated.

Methodology (cont.)

In our system, the pulses were chirped using a four-pass, single-grating, non-imaging telescopic stretcher, a simplified version of the many such systems designed by C. P. J. Barty and co-workers at Stanford University⁶. [See Fig. 3 below for a detailed sketch of our as-built stretcher.] The assembly makes use of two plane mirrors, a curved mirror with a 0.5 m radius of curvature, and a precision-ruled 1200 lines/mm diffraction grating blazed for 750 nm, with a protective gold coating. The grating provides spatial dispersion to a given pulse, spreading its various frequency components out into a horizontal line about 3" across. Propagation through the telescope assembly gives the temporal dispersion needed to broaden the pulse in time, by allowing, e.g., the red components of the pulse to pull ahead of the blue due to their different indices of refraction. Careful orientation of the stretcher components enables the input pulses to make multiple passes through the

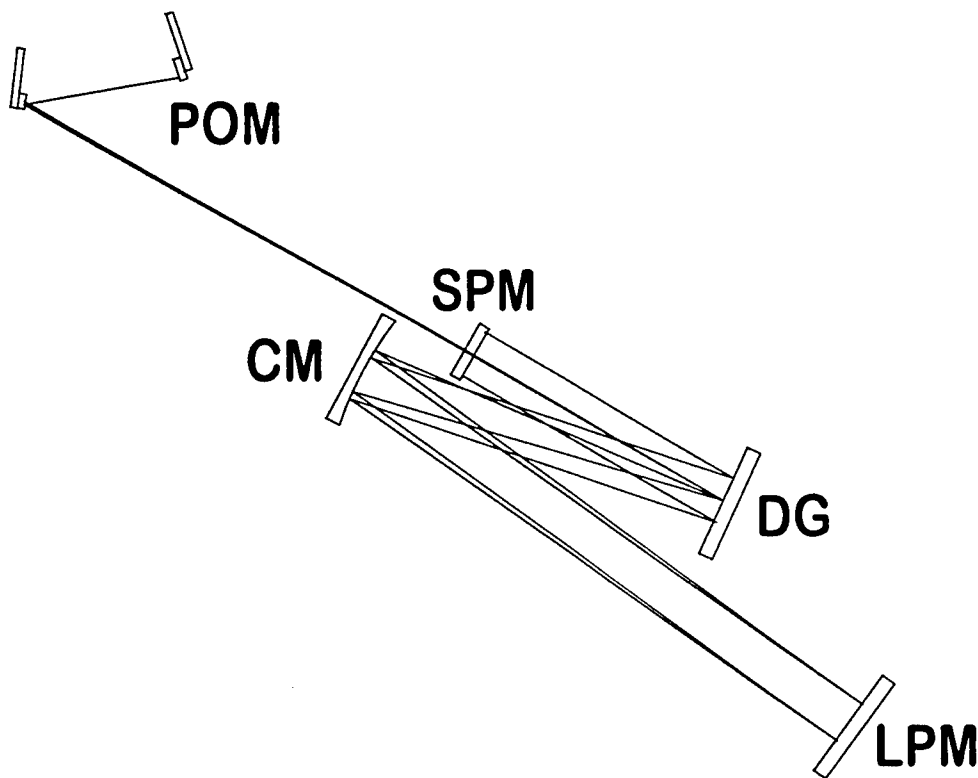


Figure 3 Detail of pulse-stretching assembly. DG -- diffraction grating; CM -- curved mirror; LPM -- large plane mirror; SPM -- small plane mirror; POM -- pick-off mirror

assembly, such that the stretched beam exits the stretcher at a different height and angle than it entered (see Fig. 3.) In particular, pulses enter the stretcher by passing above the small plane mirror (SPM in the diagram); after two roundtrips between the two plane mirrors, they exit below the smaller one. The beam's path walks downward on the mirrors due to the slight downward tilt of the large plane mirror (LPM). As a result, the exit beam is directed by a fold mirror to a pick-off mirror, instead of back into the Ti:S oscillator. The pick-off mirror in turn sends the stretched pulsetrain towards the regenerative amplifier.

Methodology (cont.)

Regenerative Amplifier Cavity

The regenerative amplification process typically consists of three basic steps: coupling the stretched pulsetrain into the regen cavity; allowing a single pulse to make multiple roundtrips through the YLF-pumped crystal so it can be amplified, then finally extracting the amplified pulse from the regen. In our case, these tasks were accomplished using a Karl Lambrecht calcite Brewster-cut prism polarizer as a polarization-sensitive input/output gate, a Medox Pockels Cell to allow electrical control over the recirculation and extraction of pulses undergoing amplification, and an EO Technology Faraday isolator to give the inbound pulses the proper polarization orientation then later route the outbound pulses to an eventual compressor stage.

In our system, the Faraday isolator is both the first and last optical component seen by the stretched pulses during the regenerative amplification process. The isolator consists of an input polarizer with its transmission axis oriented horizontally, a Faraday rotator, a quartz rotator, and an output polarizer oriented vertically. Since the mode-locked pulses generated by the oscillator are horizontally polarized, they pass through the initial polarizer unaffected. They are rotated 45 degrees by the Faraday rotator, then an additional 45 degrees by the quartz rotator. As a result, they pass through the output polarizer unaffected, emerging vertically polarized, and are directed by a pair of fold mirrors to the Brewster prism polarizer, which is located within the regen cavity. The pulses enter the polarizer through one of its input/output ports, and because of their polarization orientation and approach angle, are reflected internally by the various prism faces back in the direction of the Pockels cell, along the regen cavity axis. At this point the pulses have successfully entered the regen and they are vertically polarized (See Fig. 4.)

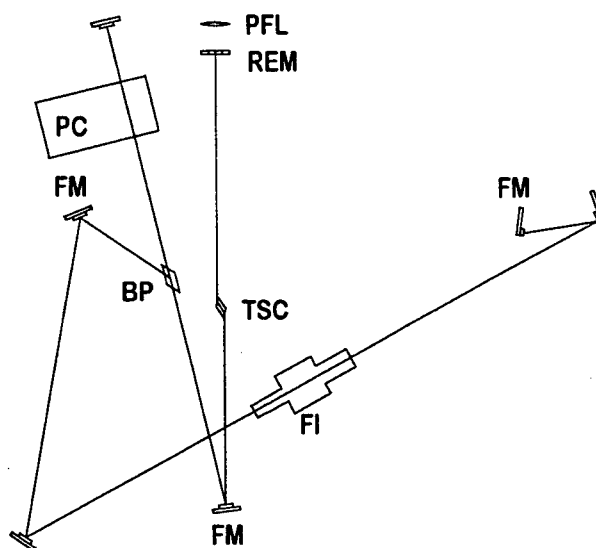


Figure 4 Regen cavity and in-coupling/extraction components. FM -- fold mirror; FI -- Faraday isolator; BP -- Brewster polarizer; PC -- Pockels cell; TSC -- Ti:Sapphire crystal; REM -- regen entrance mirror; PFL -- pump focus lens.

Methodology (cont.)

The intracavity Pockels cell acts as a voltage-controllable retarder with three different possible transmission states depending on the voltage applied to it. With no applied voltage, the cell imparts a static quarter-wave retardation to incoming light. Since pulses pass through the cell twice (once prior to reflection at the high-reflector and once as they return), they experience a net half-wave of retardation, and are rotated by 90 degrees from their previous polarization orientation. So the pulses admitted by the Brewster prism polarizer are rotated from vertical to horizontal as a result of their initial two-pass traversal of the Pockels cell.

Next they approach the Brewster polarizer once more. Since they are now horizontally polarized, they pass directly through the polarizer and complete a roundtrip through the arm of the regen cavity containing the Ti:Sapphire crystal. At this point the crystal is not being pumped, so the pulses experience no gain. On their return trip, they again pass through the Brewster polarizer and approach the Pockels cell for the second time. The pulses' polarization is flipped again, this time from horizontal to vertical, so that when they return to the Brewster polarizer, they are reflected out of the regen cavity through the same port they originally entered.

Retracing their original path, the pulses pass through the Faraday isolator in reverse, but with a different net rotation. As before, the vertically-polarized pulses pass the output polarizer unaffected, and their polarization is rotated 45 degrees by the quartz rotator. This time, however, the Faraday rotator undoes the rotation performed by the quartz rotator, rather than adding to it, and the pulses remain vertically polarized. They then exit the input polarizer through a different face of the polarizer than they entered, and are directed towards the compressor assembly (not shown in Fig. 4) instead of back towards the stretcher where they originally came from. So when no pump-energy is supplied to the regen crystal and no voltage is applied to the Pockels cell, input pulses make one roundtrip of the regen cavity and exit without being amplified.

To create a population inversion and gain within our regen cavity, we used a Q-switched mode-locked Nd:YLF pump laser⁷ to generate ~100 ns, 532 nm pulses at a 1 kHz repetition rate. During the 100 ns window that the crystal is pumped by the Nd:YLF and the cavity acts as an amplifier, the input pulses from the stretcher need to be allowed to circulate more than just once through the Ti:Sapphire crystal. This is accomplished by applying a voltage to the Pockels cell sufficient to generate a quarter-wave retardance in addition to the static quarter-wave bias, with the applied voltage being perfectly synchronized to the Q-switch frequency of the pump laser. Under these conditions, the last pulse which entered the regen prior to the quarter-wave voltage being applied (and thus left the Pockels cell horizontally polarized) will see gain and amplification as it traverses the pumped Ti:Sapphire crystal, and will see a full wave of retardance during subsequent roundtrips through the Pockels cell as long as this voltage is kept on. Because it remains horizontally polarized, this pulse will recirculate through the cavity without being switched out by the Brewster polarizer. On the other hand, vertically-polarized pulses input to the regen after the quarter-wave voltage is applied see a full wave of retardance during their initial roundtrip through the

Methodology (cont.)

Pockels cell; thus they remain vertically polarized and are switched back out by the Brewster polarizer without being amplified.

To switch the recirculating pulse out of the amplifier cavity after the desired amount of roundtrips, its polarization must be flipped from horizontal back to vertical. This is accomplished, not by turning off the voltage to the Pockels cell, but by applying an additional quarter-wave voltage step. In this fashion, the Pockels cell becomes a three-quarter wave retarder, so that two passes through it impart a net half-wave retardance to the amplified pulse, which is then switched out of the regen cavity via the Brewster polarizer. Thus the temporal duration of the first quarter-wave voltage step applied to the Pockels cell determines the number of amplification roundtrips experienced by a given pulse. For a quarter-wave voltage time window of approximately 160 ns, as we used, and a regen cavity length designed to produce a 10 ns roundtrip time, the pulse being amplified undergoes roughly 15 roundtrips, with an intensity gain of a little over two per pass, or about a factor of 10,000 amplification.

Results

After significant early effort to obtain mode-locked femtosecond pulses from our oscillator, after satisfactorily resolving our pump-laser instability problems, and after replacing our initial Ti:Sapphire crystal (which experienced significant facet damage, due to inadequate cooling) we were able to consistently observe ~65 fs pulses for as long as several hours before having to "tweak" cavity mirrors or bang the table to restore mode-locking. A typical pulse spectrum under stable conditions is shown in Fig. 5. A more frequently-observed spectrum when the oscillator was not behaving in a stable fashion, or even sporadically during otherwise stable operation, is given in Fig. 6. Here we see the effects of CW breakthrough, with the strong, narrow-linewidth CW peak superimposed on the broad mode-locked spectrum. These spectra were obtained using an Oriel InstaSpec diode array spectrometer system with roughly 1 nm resolution.

Optimization of the regenerative amplifier cavity and the Pockels cell driver settings (electrical pulse duration, delay timing, triggering, etc.) eventually yielded the desired amplification and extraction from the regen of an individual pulse at a 1 kHz repetition rate. The amplification process is readily visible in Fig. 7, where an oscilloscope trace was obtained by placing a high-speed detector and a small-transmission output coupler in place of the regen high-reflector. The pulses traverse the regen at a rate of 10 ns per roundtrip, which accounts for the separation between detected spikes. Their intensity is approximately doubled with each complete pass through the amplifier, until the Q-switched pulse from the Nd:YLF pump laser begins to decay. In this figure, approximately 18 roundtrips are visible underneath the Q-switched pulse build-up and decay envelope. Many more are not visible, however, beyond the left edge of the figure.

Results (cont.)

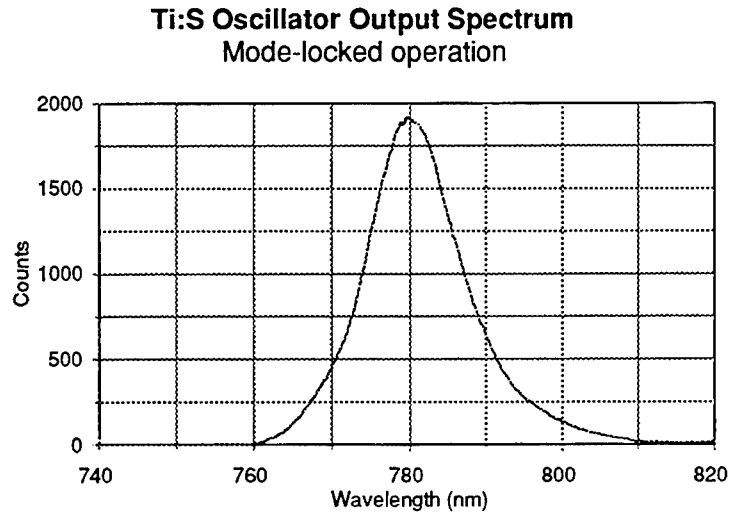


Figure 5 Typical output spectrum for mode-locked operation of our Ti:S oscillator. Assuming a sech^2 pulse, the temporal FWHM of the pulse is estimated at ~ 65 fs in this case. There is some asymmetry visible in the spectrum, and a small (~ 100 counts) satellite is barely visible at the right edge of the graph.

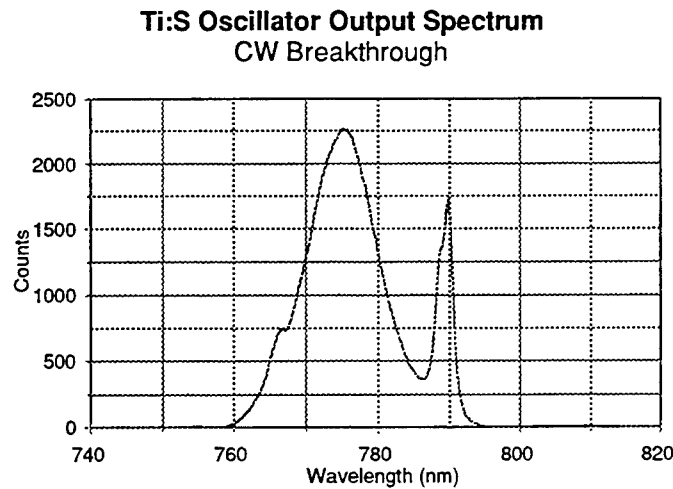


Figure 6 The CW breakthrough phenomenon, a frequently-manifested feature in our oscillator output. We attribute this behavior in our case to poor pump-laser coupling into the Ti:Sapphire crystal, incomplete mode-matching, and thermal fluctuations due to unsatisfactory spatial and temporal behavior on the part of our Argon-Ion pump laser.

Results (cont.)

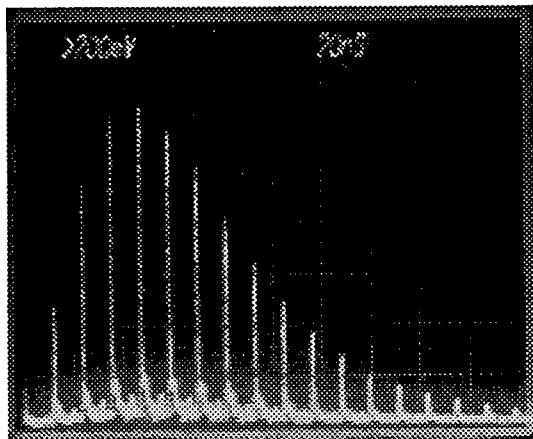


Figure 7 Buildup of a pulse undergoing amplification within the regen cavity. Note how the second and third spikes are approximately a factor of two larger than their immediate predecessors, indicating a gain factor of roughly two per roundtrip.

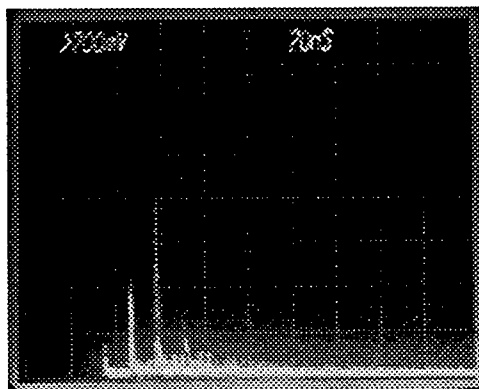


Figure 8 Dumping of the regen cavity by the Pockels cell. When the second quarter-wave voltage step is applied, the pulse undergoing amplification within the cavity is switched out by polarization rotation and by reflection from the intracavity Brewster prism. Thus the oscilloscope trace indicates an abrupt halt in the recirculation process. Note by comparison with Fig. 7 how the pulse has been switched out near its peak amplitude.

The desired result of the regenerative amplification process can be seen in Fig. 8. Here we have applied both the first and second quarter-wave voltage steps to the Pockels cell. A single pulse makes approximately 15 roundtrips within the regen cavity (only the last three are visible in the figure), then is switched out of the cavity by the second voltage step. The amplified pulse is ready to be re-compressed to femtosecond timescales, then focused onto a suitable material for continuum generation. As it turns out, the pulses shown here, even uncompressed, have sufficient peak energy to generate a weak continuum in a quartz substrate.

Conclusions

Though we were unable to complete the originally-planned optical parametric amplifier portion of our project during the time allotted, we were successful in designing and building a regenerative amplifier system capable of producing pulses with sufficient energy for continuum generation. As built, the system generated 0.5 mJ, ~100 ps pulses at a 1 kHz repetition rate, and should serve as a highly useful source of both narrowband pump and broadband probe pulses for ultrafast spectroscopy applications, including parametric amplification in the 1.3 μm region.

References

1. See, e.g., N. Peyghambarian, et al., *Introduction to Semiconductor Optics*, Englewood Cliffs, New Jersey, Prentice-Hall, 1993.
2. See, e.g., W. H. Knox et al., "Amplified femtosecond optical pulses and continuum generation at 5-kHz repetition rate," *Opt. Lett.* vol. 9, 552 (1984).
3. J. Squier, G. Korn, G. Mourou, G. Vaillancourt, and M. Bouvier, "Amplification of femtosecond pulses at 10-kHz repetition rates in Ti:Al₂O₃," *Opt. Lett.* vol. 18, 625 (1993).
4. M. T. Asaki et al., "Generation of 11-fs pulses from a modelocked Ti:sapphire laser," *Opt. Lett.* vol. 18, 977 (1993).
5. D. Strickland and G. Mourou, "Compression of amplified chirped optical pulses," *Opt. Comm.* vol. 56, 219 (1985).
6. B. E. Lemoff and C. P. J. Barty, "Quintic-phase-limited, spatially uniform expansion and recompression of ultrashort optical pulses," *Opt. Lett.* vol. 18, 1651 (1993).
7. G. Vaillancourt, T. B. Norris, J. S. Coe, P. Bado, and G. A. Mourou, "Operation of a 1 kilohertz pulse-pumped Ti:Sapphire regenerative amplifier," *Opt. Lett.* vol. 15, 317 (1990).

A STUDY OF THE SYNTHESIS AND PROPERTIES
OF ORGANICALLY-MODIFIED METAL ALKOXIDES
FOR OPTICAL AND ELECTROOPTICAL THIN-FILM APPLICATIONS

Daniel F. Ryder, Jr.
Associate Professor
Department of Chemical Engineering

Tufts University
Department of Chemical Engineering
4 Colby Street
Medford, Massachusetts 02155

Final Report For:

Summer Faculty Research Program
Rome Laboratory

Sponsored by:

Air Force Office of Scientific Research
Bolling Air Force Base, DC

and

Rome Laboratory

December 1995

A STUDY OF THE SYNTHESIS AND PROPERTIES
OF ORGANICALLY-MODIFIED METAL ALKOXIDES
FOR OPTICAL AND ELECTROOPTICAL THIN-FILM APPLICATIONS

Daniel F. Ryder, Jr.
Associate Professor
Department of Chemical Engineering
Tufts University

Abstract

The development of materials processing technology for optical and electrooptical thin films remains a principle research area of electronic materials science. The application of chemical processing techniques, such as the sol-gel and MOCVD methods, has driven developmental research relating to chemical precursor chemistry. In this work, the synthetic chemistry and properties of organically-modified metal alkoxides were studied from two specific application perspectives of current technological interest : 1.) volatile, single-source, complexed metal compounds for MOCVD applications, and 2.) durable, antireflective coating systems for optical polymers.

In Part 1 of this work, the synthesis and properties of β -diketonate-modified heterobimetallic alkoxides was examined as an approach to formulate a volatile, single-source, complexed metal compound for the subsequent CVD processing of BaTiO_3 . While synthetic approaches for the formation of a $\text{BaTi}(\text{isopropoxide})_{n-x}(\text{2,2,6,6-tetramethyl-3,5-heptanedionate})_x$ were developed, subsequent sublimation trials showed that thermally induced decomposition reactions resulted in nonstoichiometric vaporization. Nevertheless, preliminary results show that the synthesized compounds may be purified by conventional techniques, and as such, these compounds may be utilized in the development of aerosol-based source delivery systems.

In Part 2 of this work, sol-gel processing methods were utilized to deposit ORganically-Modified CERamic (ORMOCer) thin films of defined refractive index. It was further demonstrated, that by utilizing standard antireflective (AR) film designs, multilayered ORMOCer thin films deposited at low temperature could form the basis of a durable AR coating system for the optical polymers (e.g., PMMA, polycarbonate, allyl diglycol carbonate). As an illustrative example, the design of a single-layer AR (i.e., SLAR) and bilayer V-type and W-type AR coatings for polycarbonate substrates were reviewed.

A STUDY OF THE SYNTHESIS AND PROPERTIES
OF ORGANICALLY-MODIFIED METAL ALKOXIDES
FOR OPTICAL AND ELECTROOPTICAL THIN-FILM APPLICATIONS

Daniel F. Ryder, Jr.

Part 1: Synthesis and Properties β -Diketonate-Modified Heterobimetallic Alkoxides

Introduction

Metalorganic chemical vapor deposition (MOCVD) of multicomponent oxides has recently been the subject of extensive research study, particularly as related to the film processing of the cuprate superconductors (e.g., $\text{YBa}_2\text{Cu}_3\text{O}_x$, $\text{Bi}_2\text{Sr}_2\text{Ca}_2\text{Cu}_3\text{O}_y$) and the ferroelectric perovskites (e.g., BaTiO_3 , $(\text{Ba,Sr})\text{TiO}_3$). Current commercial practice involves the controlled vaporization and transport of individual metal precursor compounds to achieve stoichiometric compositional control of the growing film. As an alternative to the complicated operational strategies associated with conventional multi-source MOCVD processing, the use of a single-source complexed metal compound which inherently possesses the desired stoichiometric ratio of the targeted oxide has been proposed. Chour *et al.*¹ recently reported on the vapor deposition of LiTaO_3 from a single-source, volatile double metal alkoxide ($\text{LiTa}(\text{O}^i\text{Bu})_6$). Likewise, Zhang *et al.*² reported on the CVD growth of MgAl_2O_4 thin films from a flash-evaporated $\text{MgAl}_2(\text{O}^i\text{Pr})_8$ precursor. A review of the definitive text Metal Alkoxides³ shows that there are a number of double metal alkoxides which possess the three properties required for a MOCVD single-source compound:

- 1) A stoichiometric composition consistent with the targeted product oxide.
- 2) Good thermal stability. Solid/liquid-phase thermal decomposition reactions to the individual metal alkoxides and/or other polymeric oxo-alkoxide intermediates would lead to nonstoichiometric deposition.

¹ K. Chour, G. Wang, and R. Xu, "Vapor Deposition of Lithium Tantalate with Volatile Double Alkoxide Precursors", in *Mat. Res. Soc. Symp. Proc. Vol. 335, Metal-Organic Chemical Vapor Deposition of Electronic Ceramics*, S.B. Desu, D.B. Beach, B.W. Wessels, and S. Gokoglu, Eds., Mat. Res. Soc., Pittsburgh, PA 65, (1994).

² J. Zhang, G.T. Stauff, R. Gardiner, P. Van Buskirk, and J. Steinbeck, "Single Molecular Precursor Metal-Organic chemical Vapor Deposition of MgAl_2O_4 thin films", *J. Mater. Res.*, 9[6], 1333, (1994).

³ D.C. Bradley, R.C. Mehrota, and D.P. Gaur, Metal Alkoxides, Academic Press, Orlando, FL, (1978).

3) Sufficient volatility at normal processing conditions.

It is notable that no bimetallic alkoxide which contains an alkaline-earth (i.e., Ba, Sr, Ca) : Ti pair possesses adequate thermal stability for CVD applications. Furthermore, and due to the propensity of the alkaline-earths to attain high coordination numbers, all of the referenced alkaline-earth-based double metal alkoxides possess metallic stoichiometric ratios inconsistent with commercially significant oxide compositions. Given the current interest in thin film ferroelectric materials based on BaTiO₃, it was proposed to experimentally investigate methods whereby the chemical composition, thermal stability, and volatility of Ba:Ti alkoxide-based systems may be modified/enhanced. In this initial study, the synthetic chemistry and properties of β -diketonate-modified Ba:Ti 2-isopropoxides were investigated. As the basis for this choice, a brief summary relating to the development of Group II source compounds is presented. In addition, background information relating to the synthesis of homo- and hetero-metal alkoxides and the reaction kinetics of metal alkoxides with β -diketonates is briefly surveyed.

Precursor Development of Group II Source Compounds for CVD Applications

In order to ensure reasonable epitaxial growth rates at nominal substrate temperatures, source compounds should transport in the vapor phase at processing temperatures less than 200 °C and pressures greater than 5 torr.⁴ This constraint presents a significant problem for the case of the alkaline earth source compounds as their propensity to attain high coordination numbers, usually by forming multinuclear aggregates, severely limits their volatility.⁵ As such, only those alkaline earth compounds with multidentate, sterically encumbered anions have proven useful for MOCVD applications.

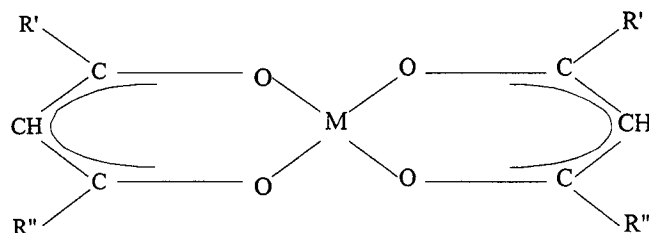
Recent synthetic chemical research has focused on both the fluorinated and fluorine-free β -diketonates which have the general structural form as shown below.^{5,6,7}

⁴ G.B. Stringfellow, Organometallic Vapor-Phase Epitaxy: Theory and Practice, Academic press, New York, (1989).

⁵ B.A. Vaartstra, R.A. Gardiner, D.C. Gordon, R.L. Ostrander, and A.L. Rheingold, "Advances in Precursor Development for CVD of Barium-Containing Materials", in Mar. Res. Soc. Symp. Proc. Vol. 335, Metal-Organic Chemical Vapor Deposition of Electronic Ceramics, S.B. Desu, D.B. Beach, B.W. Wessels, and S. Gokoglu, Eds., Mat. Res. Soc., Pittsburgh, PA 203, (1994).

⁶ H.A. Meinema, K. Timmer, H.L. Linden, and C.I.M.A. Spee, "Synthesis Strategies for MOCVD Precursors for HTcS Thin films", in Mar. Res. Soc. Symp. Proc. Vol. 335, Metal-Organic Chemical Vapor Deposition of Electronic Ceramics, S.B. Desu, D.B. Beach, B.W. Wessels, and S. Gokoglu, Eds., Mat. Res. Soc., Pittsburgh, PA 193, (1994).

⁷ W.S. Rees Jr., H.A. Luten, M.W. Carris, C.R. Caballero, W. Hesse, and V.L. Goedken, "Current Status of Recent Results on Group 2 Source Compounds for Vapor Phase Epitaxy of Ferroelectric Thin Films", in Mat. Res. Soc. Symp. Proc. Vol. 310, Ferroelectric Thin Films III, E.R. Myers, B.A. Tuttle, S. Desu, and P.K. Larsen, Eds., Mat. Res. Soc., Pittsburgh, PA, 375, (1993).



As illustrated, a monomeric alkaline earth β -diketonate would be coordinately unsaturated with a coordination number (CN) of 4, where a CN of 6 to 10 is the thermodynamically-stabilized structural form. As such, intermolecular association takes place resulting in the formation of molecular clusters which adversely effects the volatility. Volatility has been reported to increase with increased encapsulation of the metal ion through the utilization of bulky side-group ligands.⁵ Additionally, the use of fluorinated alkyl groups results in a decrease of intermolecular interactions which consequently increases volatility. For the case of fluorinated precursors, the resulting enhancement in volatility is at the expense of a more complicated decomposition reaction pathway to the oxide through the hydrolysis of a thermally stable fluoride salt. As such, more recent studies have concentrated on the non-fluorinated source compounds such as those based on 2,2,6,6-tetramethyl-3,5-heptanedionate (thd).

$\text{Ba}(\text{thd})_2$ forms a tetrameric cluster in the solid state which is maintained in the transformation to the gas phase.^{5,8} While $\text{Ba}(\text{thd})_2$ is air stable, significant decomposition through hydrolysis and carbonate formation precludes synthesis through the standard "wet" method⁹. Pure $\text{Ba}(\text{thd})_2$ may be synthesized by direct reaction of the metal or metal hydride with two (2) equivalents of the free β -diketone (Hthd). Based on nominal growth rate requirements, use of pure $\text{Ba}(\text{thd})_2$ as a barium precursor requires sublimation temperatures in slight excess of 200 °C. Prolonged heating at these temperatures induces thermal decomposition processes which result in the significant deterioration of the vapor transport rate. Methods to induce dissociation of the tetrameric $\text{Ba}(\text{thd})_2$ clusters to monomeric species would serve to enhance volatility, and various methods to induce coordinative saturation through adduct addition have been reported. For example, evaporation of $\text{Ba}(\text{thd})_2$ in the presence of oxygen (e.g., polyethers) or nitrogen (e.g., THF, NH_3 , $\text{N}(\text{CH}_3)_3$) donor ligands has been reported to enhance the thermal stability and the volatility of the compound such that increased mass transport rates may be maintained at lower evaporation temperatures.⁶

⁸ J.E. Schwarberg, R.e., Sievers, and R.W. Moshier, "Gas Chromatographic and Related Properties of the Alkaline Earth Chelates with 2,2,6,6-Tetramethyl-3,5-Heptanedione", *Anal. Chem.*, **42**(14), 1828, (1970).

⁹ K.J. Eisentraut and R.E. Sievers, "Volatile Rare Earth Chelates", *J. Am. Chem. Soc.*, **87**(22), 5254, (1965).

Synthesis and Properties of Homo- and Hetero-Metal Alkoxides

The reader is referred to the text by Bradley *et al*⁸ for general information regarding the synthesis and properties of metal alkoxides, as the background information included herein is specific to the alkoxides of the alkaline earths and the heterobimetallic alkoxides of the alkaline earths and titanium.

The alkaline earth alkoxides may be synthesized via the direct reaction of the metal and/or metal hydride with the alcohol. Simple $[M(OR)_2]_m$ ($R = \text{Me, Et, Pr}^n, \text{Pr}^i$) alkoxides of Group 2 metals are generally polymeric, insoluble in organic solvents, and non-volatile. Lowering the degree of association has been shown to increase the solubility and volatility, and the role of steric factors on the extent of oligomerization has been actively studied. Recent strategies reported in the literature to induce dissociation include the utilization of sterically demanding ligands (e.g., $2,6\text{-Bu}_2\text{C}_6\text{H}_3\text{O}^-$, Et_3CO^- , $\text{C}(\text{Me}_3)_2\text{CHO}^-$, etc.) and chelating ligands (e.g., $\text{Me}_2\text{NCH}_2\text{CH}_2\text{O}^-$, $\text{ROCH}_2\text{CH}_2\text{O}^-$, $\text{CH}_3(\text{OCH}_2\text{CH}_2)_n\text{O}^-$, etc.).¹⁰ Relative to the 2-methoxyethoxide chelating ligand, it is important to note that there is an additional possibility for this ligand to act as a bridging ligand which may lead to the formation of large oligomers.¹¹ This property has been the basis for the use of 2-methoxyethoxide as an assembling ligand for the formation of heterobimetallic alkoxides.

As compared to associated homo-metallic alkoxides, the heterobimetallic alkoxides exhibit, in general, and enhance stability probably resulting from the greater stability of the $M(m\text{-OR})_2M'$ bridges between the two dissimilar metals (i.e., M and M'). In addition, a number of heterobimetallic alkoxides containing an alkaline earth metal constituent are soluble, volatile, and monomeric - which contrasts sharply with the insoluble, polymeric, and non-volatile nature of the associated homo-metallic alkoxides. Bimetallic alkoxides containing an alkaline earth constituent have been synthesized by the Lewis acid-base interaction of the simple alkoxide of these metals with the alkoxides of less electropositive elements. D.A. Payne *et al*¹² recently reported on the synthesis of barium titanium methoxyethoxide $[\text{BaTi}(\text{OCH}_2\text{CH}_2\text{OCH}_3)_6]$ via this method. In previous work¹³, the structure of similar lead titanium alkoxide was reported to be consistent with that of a linear polymeric network with an empirical chemical formula of $\text{PbTi}(\text{OCH}_2\text{CH}_2\text{OCH}_3)_6$. As such, it may be concluded that chelating ligands, such as 2-methoxyethoxide, not only act as bridging ligands conducive to the formation of heterobimetallic

¹⁰ R.C. Mehrotra, A. Singh, and S. Sogani, "Homo- and Hetero-metallic Alkoxides of Group 1, 2, and 12 Metals", *Chem. Soc. Rev.*, pp. 215-25, (1994).

¹¹ L.G. Hubert-Pfalzgraf, O. Poncelet, C. Sirio, "Tailoring Metal Alkoxides Using Functional Alcohols: Some Examples in Yttrium, Copper, and Main-Group Chemistry (Bismuth, Barium)" in *Chemical Processing of Advanced Materials*, L.L. Hench and J.K. West, Eds., John Wiley & Sons, Inc., New York, NY, 277, (1992).

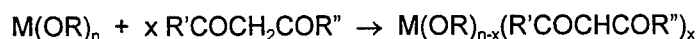
¹² D.A. Payne, D.J. Eichorst, L.F. Francis, and J-F. Campion, "Molecular Precursors for the chemical Processing of Advanced Electrical Ceramics" in *Chemical Processing of Advanced Materials*, L.L. Hench and J.K. West, Eds., John Wiley & Sons, Inc., New York, NY, 499, (1992).

¹³ S.D. Ramamurthi and D.A. Payne, "Structural Investigations of prehydrolyzed Precursors Used in the Sol-Gel Processing of Lead Titanate", *J. Am. Ceram. Soc.*, **73**[8], 2547, (1990).

alkoxides, but also induce the formation of large oligomers. The resulting polymeric structures are, in general, nonvolatile and as such bimetallic 2-methoxyethoxides are not suitable source compounds for CVD applications.

Chemical Reactions of Metal Alkoxides with β -Diketone

The enolic form of β -diketone contains a reactive hydroxyl group and this molecule reacts very readily with metal alkoxides, which may be represented by the following general reaction scheme:



where, M	Metal
R, R', R''	Alkyl Group
n, x	stoichiometric coefficients

The specific reaction chemistry of titanium (IV) alkoxides with compounds having bidentate ligands (e.g., acetylacetone) has been well studied and the reader is referred to the literature^{14,15} for details beyond those presented herein. In general, the reactions of titanium alkoxides (i.e., methoxide, ethoxide, isopropoxide) with acetylacetone (Hacac) have been found to yield two types of monomeric product derivatives: a penta-coordinated titanium complex ($Ti(OR)_3(acac)$) when a 1:1 ratio of titanium (IV) alkoxide to acetylacetone reactant is used, and a hexa-coordinated titanium complex ($Ti(OR)_2(acac)_2$) with a 1:2 ratio of titanium (IV) alkoxide to acetylacetone. In addition, Yamamoto and Kambara¹⁶ have noted that even when titanium alkoxides are treated with an excess of acetylacetone under normal conditions, only titanium dialkoxy diacetylacetone monomeric products are formed. Although hexa-coordinated, it was also reported that the structure is sufficiently labile to coordinate and react with higher alcohols.

¹⁴ D.M. Puri and R.C. Mehrotra, "Derivatives of Titanium and Compounds Having Bidentate Ligands II: Reactions of Titanium Alkoxides with Methyl Acetoacetate", *J. Less-Common Met.*, **3**, 253, (1961).

¹⁵ D.M. Puri and R.C. Mehrotra, "Derivatives of Titanium and Compounds Having Bidentate Ligands III: Reactions of Titanium Alkoxides with Acetylacetone", *J. Less-Common Met.*, **4**, 393, (1962).

¹⁶ A. Yamamoto and S. Kambara, "Structures and Reaction Products of Tetraalkoxytitanium and Acetylacetone and Ethyl Acetoacetate", *J. Am. Chem. Soc.*, **74**, 4344, (1957).

Experimental

Synthesis

2,2,6,6-Tetramethyl-3,5-heptanedione (95%) (Hthd), titanium (IV) isopropoxide, and 2-methoxyethanol (anhydrous) were obtained from Aldrich Chemical Company and used as delivered. Barium isopropoxide used in this study was either obtained commercially from Alfa Chemical Company or synthesized directly via the reaction of the metal with isopropanol. Benzene and isopropanol solvents used in this study were dried over CaH_2 and purified by distillation. All synthesis chemistry was conducted in a controlled atmosphere glovebox.

Barium isopropoxide was synthesized via standard methods.¹⁷ Barium metal shot (3.37 g) was dispersed in 50 ml. of anhydrous benzene. Dry isopropanol was slowly added to this dispersion while stirring. After addition of the isopropanol, the solution was atmospherically refluxed ($T \approx 80^\circ\text{C}$) for four hours. The cooled, product solution was filtered off and stored for further processing. In some cases, a commercially available $\text{Ba}(\text{O}^i\text{Pr})_2$ was used as purchased and diluted with dried benzene to targeted concentrations.

The reaction of both titanium isopropoxide and barium isopropoxide in benzene with 2,2,6,6-tetramethyl-3,5-heptanedione (Hthd) was easily accomplished at room temperature. For the case of the barium compound, comparison of the thermogravimetric analysis of a vacuum dried compound sample to that of commercially available $\text{Ba}(\text{O}^i\text{Pr})_2$ and $\text{Ba}(\text{thd})_2$ showed that the synthesized compound exhibited oxidative decomposition kinetics consistent with $\text{Ba}(\text{thd})_2$.

Various methods to synthesize a $\text{BaTi}(\text{O}^i\text{Pr})_{n-x}(\text{thd})_x$ compound were experimentally studied. While quantitative analysis of product purity and structure is lacking, the preferred method resulted in the formation of a uniform yellowish-white powder product (BT #1) which was observed to uniformly melt to a straw-yellow liquid during subsequent sublimation trials. The synthesis procedure utilized for the production of product BT #1 is outlined below:

Procedure for Synthesis of Product BT #1

1. All solvents (i.e., benzene and isopropanol) were dried over CaH_2 and purified by distillation prior to use.
2. Prepare $\text{Ti}(\text{O}^i\text{Pr})_2(\text{thd})_2$
 - To 250 ml. three-neck flask reactor equipped for atmospheric reflux, add:
 - 5 g titanium isopropoxide (97%)
 - 6.48 g 2,2,6,6-Tetramethyl-3,5-heptanedione (95%)
 - 50 g benzene (thiophene free)
 - Atmospheric reflux ($T \approx 80^\circ\text{C}$) for 3 hours
 - Cool to room temperature (Prod #1)

¹⁷ H. Okamura and H.K. Bowen, "Preparation of Alkoxides for the Synthesis of Ceramics", *Ceramics International*, **12**, 161, (1986).

3. Prepare Ba(OiPr) (thd)
 - To 250 ml. three-neck flask reactor equipped for atmospheric reflux, add:
5.3 g Ba (metal chips)
59.3 g isopropanol
 - Mix at room temperature for ~ 15 hours (i.e., overnight)
 - Add 35.5 g benzene (thiophene free)
 - Add 7.11 g 2,2,6,6-Tetramethyl-3,5-heptanedione (95%)
 - Atmospheric reflux ($T \approx 80^{\circ}\text{C}$) for 3 hours
 - Cool to room temperature (Prod #2)
4. Prepare BaTi(OiPr)_{n-x}(thd)_x
 - To Prod #1 solution, add 48.87 g of Prod #2 solution.
 - Atmospheric reflux ($T \approx 80^{\circ}\text{C}$) for 3 hours
 - Cool to room temperature
5. Remove solvents by room-temperature vacuum distillation

Sublimation Studies

Sublimation trials were conducted in a standard laboratory glass sublimator. A mechanical pump was used to maintain vacuum pressure, and a hot plate was used to control the vaporization temperature. Temperature was indicated by a thermocouple located on the bottom outside wall of the sublimator. The sublimator cold-finger was cooled with an acetone-dry ice bath. In general, experiments were conducted under full mechanical pump vacuum and at processing temperatures between 200-240 °C.

Runs were terminated when sufficient materials were collected on the sublimator cold finger. Sem/EDS methods were utilized to characterize the Ba:Ti composition of representative samples.

Results

SEM/EDS analysis was used as the primary means of measuring Ba:Ti ratio of both the synthesized materials and sublimation trial residues. The dried Ba-Ti compound had an overall atomic composition of 52/48 Ba:Ti which is well within the experimental error for a 1/1 ratio. The measured metals composition was consistent over both small and large analysis areas indicating the material was fairly homogeneous. In addition, there were no microstructural features which would indicate a multiphasic composition. Consistent with these findings, the residue remaining after an oxidative, thermogravimetric analysis was found to have a 48/52 Ba:Ti.

After batch sublimation trials, samples were taken from the cold finger and the bottom of the reaction flask for analysis. In all cases, the cold finger (i.e., volatile) product was found to be barium deficient ranging from a 7/93 to a 22/78 Ba:Ti atomic ratio. Likewise, and consistent with these findings, the flask bottom composition was always barium rich averaging a 70/30 Ba:Ti atomic ratio. As such, we were not successful in developing a volatile Ba:Ti organometallic via the reviewed approach. It is not clear from the data analyzed whether a true Ba:Ti compound was formed which subsequently underwent a thermal decomposition to its individual metal constituents or if the synthesis approach led to the formation

of a solution of individual metal components. One notable result of this work was that the formed material was soluble in common alcohols (i.e., isopropanol) which is sometimes problematic with alkaline earth alkoxides.

Conclusion

In this work, the synthesis and properties of β -diketonate-modified heterobimetallic alkoxides was examined as an approach to formulate a volatile, single-source, complexed metal compound for the subsequent CVD processing of BaTiO_3 . While synthetic approaches for the formation of a $\text{BaTi}(\text{isopropoxide})_{n-x}(\text{2,2,6,6-tetramethyl-3,5-heptanedionate})_x$ were developed, subsequent sublimation trials showed that thermally induced decomposition reactions resulted in nonstoichiometric vaporization. Nevertheless, preliminary results show that the synthesized compounds may be purified by conventional techniques, and as such, these compounds may be utilized in the development of aerosol-based source delivery systems.

Part 2: Application of ORMOCer thin films as Durable, Anti-Reflective Coatings on Optical Polymers

Introduction

Coated thermoplastic polymers have replaced glass in a number of applications where moldability, toughness, and reduced weight have been selected as key performance properties. The primary limitations to the use of optical polymers relate to their poor surface resistance to abrasion, chemicals, and environmental factors (e.g., UV-induced degradation)¹⁸. As such, research relating to performance coatings for the optical polymers has concentrated on the development of protective coatings. While a wide host of coating systems have been reported in the technical and patent literature, nearly all of these primary coatings may be classified into one of three groups: organic (e.g., uv-cured polyacryloylated polyol¹⁹), organo-siloxane, and silica-filled organo-siloxane composites²⁰. As illustrated in Table 1, a

Table 1
Some Properties of Commercially Available Optical Plastics

Plastic	Supplier	Density (g/cm ³)	Coeff. Of expansion (10 ⁻⁶ /°C)	Upper limit of stability (°C)	n _D (20°C)	V-value
Allyl diglycol carbonate	"CR-39" PPG	1.32	90-100	60-70	1.498	53.6
Polymethyl- methacrylate	"Plexiglass" Rohm&Haas	1.19	63	70-100	1.492	57.8
Copolymer styrene- methacrylate	"Zerlon" Dow Chemical	1.14	66	95	1.533	42.4
Polycarbonate	"Lexan" GE	1.2	70	120-135	1.586	29.9
Copolymer styrene- acrylonitrile	"Lustran" Monsanto	1.07	70	90	1.569	35.7

¹⁸ W.C. Harbison, "Coatings for Plastic Glazing", *Automotive Engineering*, pp. 24-28, May, 1993.

¹⁹ U.S. Patent No. 5,459, 176 Soane Technologies, (1995).

²⁰ U.S. Patent No. 4,355,135 "Tintable Abrasion Resistant Coatings", J.R. January, Dow Corning Corp., (1982).

number of performance factors will influence the selection of a particular polymer for a given application. For example, and referring to the F/A-18 Windshield Environmental Endurance Verification Tests manual Section 4.2.6 Operational Temperature and Aerodynamic Heating, it states that the windshield design must be able to withstand a temperature cycling pattern which reaches a maximum temperature of 104 °C on “Hot Day” profiles. As such, only polycarbonate and a recently developed imidized - polymethylmethacrylate (not shown) would meet this operating temperature requirement. In general, the mechanical strength of polycarbonates is superior to the polymethylmethacrylates, but their higher refractive index would result in an unacceptable increase in surface glare. Similar to optical glass applications where maximum transmission is a design parameter (e.g., solar thermal and photovoltaic applications), antireflection (AR) coatings may be utilized to improve the optical properties. While thin film AR design formulas are readily available in the literature, materials suitable for application on temperature-sensitive polymers have not been adequately developed. Attempts at utilizing AR materials technology developed for glass systems have not been successful, as the films are not sufficiently adherent nor durable^{21,22}. In this work, sol-gel processing methods were utilized to deposit ORGanically-Modified CERamic (ORMOCer) thin films of defined refractive index. It was further demonstrated, that by utilizing standard antireflective (AR) film designs, multilayered ORMOCer thin films deposited at low temperature could form the basis of a durable AR coating system for the optical polymers (e.g., PMMA, polycarbonate, allyl diglycol carbonate). As an illustrative example, the design of a single-layer AR (i.e., SLAR) and bilayer V-type and W-type AR coatings for polycarbonate substrates were reviewed.

Processing and Properties of ORMOCer Materials

The terms ORMOSil (i.e., organically modified silicate) and, the more general, ORMOCer (i.e., organically modified ceramic) were coined by H. Schmidt^{23, 24, 25} to describe materials, processed via a sol-gel method, which include a metal-carbon bound organic radical into the glass network. The principle example found in the literature being an epoxide substituted alkoxy silane-based formulation used in a number of optical applications ranging from a protective, abrasion resistant film^{24, 25} for polymers to a new

²¹ C.S. Ashley and S.T. Reed, “Sol-Gel Films for Solar Applications” in Mar. Res. Soc. Symp. Proc. Vol. 73, Better Ceramics Through Chemistry II, Mat. Res. Soc., Pittsburgh, PA, pp. 671-77, (1986).

²² R.B. Pettit, C.S. Ashley, S.T. Reed, and C.J. Brinker, “Antireflection Films from the Sol-Gel Process”, in Sol-Gel Technology, L. Klein, Ed., Noyes Publ., pp 80-109, (1994).

²³ H. Schmidt and H. Wolter, “Organically Modified Ceramics and Their Applications”, *J. Non-Cryst. Solids*, **121**, 428, (1990).

²⁴ H. Schmidt, “Organically Modified Ceramics - Materials with “History” or “Future””, in Ultrastructure Processing of Advanced Materials, D.R. Uhlmann and D.R. Ulrich, Eds., John Wiley & Sons, NY, pp 410-423, (1993).

²⁵ H. Schmidt and B. Seiferling, “Chemistry and Applications of Inorganic-Organic Polymers (Organically Modified Silicates)”, in Mar. Res. Soc. Symp. Proc. Vol. 73, Better Ceramics Through Chemistry II, Mat. Res. Soc., Pittsburgh, PA, pp. 739-50, (1986).

bulk material for contact lenses²⁶. In general, ORMOCers are thermally cured to a dense material at temperatures ranging from 80 - 130 °C, and hence are suitable for applications as performance coatings on the polycarbonates and imide-modified polymethacrylates. In addition, specific ORMOCer formulations modified by acrylic functional groups may be uv-cured, and hence may be applied to nearly all polymeric materials. As previously noted, a majority of the patent literature on abrasion resistant coating formulations for polymers would be classified by Schmidt as silica-filled, ORMOSil-based systems.

Given the reported durability of ORMOCer materials and the capability to tailor their refractive index through the judicious choice of both the metal oxide network and organic modifier composition²⁷, the use of an ORMOCer-based AR coating was contemplated. The index of refraction of hypothetical ORMOCer formulations may be estimated by the Lorentz-Lorentz calculation for molar refraction. In addition, experience dictates that the mechanical properties of ORMOCer coatings utilized for polymer applications are optimized when the composition of functional elements are within the guidelines presented in Table 2. As an example, a target low index and high index material formulation is presented. These

Table 2
ORMOCer Formulations for AR Coatings

	wt.%	Low-Index System	High-Index System
Inorganic / Organic Network Former	25-50	Epoxysilane	Epoxysilane
Inorganic Network Former	30-65	Aluminum Alkoxide	Titanium Alkoxide
Organic Network Former		-	-
Inorganic Network Modified	10-20	Methyl-trialkoxysilane	Phenyl-trialkoxysilane

noted formulations provided the basis of the subsequent experimental study. As noted in Table 2, the properties of ORMOCer materials may be tailored through the design of their four (4) primary building blocks. The Inorganic/Organic Network Former serves as the metal-carbon linkage point for the nanocomposite system.. The Inorganic Network Former, as its name implies, determines the characteristics of the "glass-like" portion of the composite. In this work, the composition of the Inorganic Network Former was primarily utilized to effect the optical properties (i.e., refractive index) of the composite. Inclusion of Organic Network Formers result in composite systems whose properties approach that of plastics. At the extreme, nanocomposite systems similar to ceramic-filled silicone rubbers may be

²⁶ G. Philipp and H. Schmidt, "New Materials for Contact Lenses Prepared from Si- and Ti- Alkoxides by the Sol-Gel Process", *J. Non-Cryst. Solids*, **63**, 283, (1984).

²⁷ B. Linter, N. Arfsten, H. Dislich, H. Schmidt, G. Philipp, and B. Seiferling, "A First Look at the Optical Properties of ORMOSils", *J. Non-Cryst. Solids*, **100**, 378, (1988).

achieved. Given that abrasion resistance is usually associated with surface hardness, the use of Organic Network Formers was not studied in this work. Inorganic and Organic Network Modifiers are utilized as polymeric chain-length modifiers. As such, they have a direct impact on the resulting mechanical properties of the composite coating. In this work, the choice of the alkyl-group on the Modifier molecule was also specifically chosen as to its impact on the optical properties of the resulting composite. Lorentz-Lorentz molar refraction calculations were utilized to design experimental system formulations.

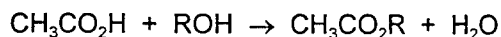
Experimental

Synthesis

The synthesis of ORMOCer coating solutions was based on a two-step modified sol-gel approach. A low-index, alumina-silicate (code: EPS-Al) formulation and a high-index titania-silicate (code: EPS-Ti) formulation were synthesized using an identical procedure. For illustrative purposes, the processing details for the alumina-silicate system are presented.

Aluminum *sec*-butoxide, triethoxymethylsilane, and γ -glycidoxypolytrimethoxysilane (i.e., epoxysilane) were obtained from Alfa Chemical Company. Anhydrous 2-propanol and ACS reagent grade hydrochloric acid (37%) were obtained from Aldrich Chemical Company. All chemicals were used as delivered. All materials transfer of alkoxide starting compounds was conducted in a controlled atmosphere glovebox.

The sol-gel method involves the hydrolysis and subsequent polycondensation of metal alkoxides to form three-dimensional polymer networks. For the case of multicomponent systems, and due to the inherent differences in reaction rates (both inter- and intra- molecular, since metal alkoxides are typically polyfunctional), the maintenance of system homogeneity is difficult. One experimental technique which has been successfully applied to a wide-host of formulations is to introduce hydrolysis water at a very slow (i.e., molecular) addition rate. As such, it has been theorized that this would minimize early self-condensation of highly reactive alkoxides (e.g., titanium and aluminum alkoxides) due to statistical kinetic effects. In practice, water may be “molecularly added” to the reaction via an *in situ* synthesis reaction or through a controlled diffusion process. The reaction of acetic acid with an alcohol ($C_nH_{2n+1}OH$ with $n > 1$,



where, R = alkyl group (e.g., C_2H_5 , C_3H_7)

most alkoxides are insoluble in methanol) may be utilized to generate reactant water *in situ*. The limitation to this method relates to the inability to remove the by-product alkyl acetate during the low temperature cure cycles associated with ORMOCer processing. Alternatively, and as previously noted, the addition rate of reactant water may be controlled by a secondary diffusion-limited process. In practice, this may be accomplished by adsorbing the reaction water onto silica gel prior to its addition to the process

reaction flask. The slow diffusion of water from the internal pores of the silica gel to the essentially anhydrous reaction mixture may be visually monitored by using the indicating silica gel commonly used for desiccants.

The low refractive index coating solution was prepared by combining a 50:30:20 (molar) ratio of γ -glycidoxypolytrimethoxysilane, triethoxymethylsilane, and aluminum *sec*-butoxide into an Erlenmeyer flask. To this mixture, a "wetted silica gel" which contained one-eighth ($1/8$) of the theoretical water necessary for complete hydrolysis and sufficient hydrochloric acid to result in a final 0.1M HCl solution was added with mixing. The resulting suspension was stirred at room temperature for approximately 3 hours. For the cases where an indicating silica gel was utilized, the progress of the reaction could be visually monitored by the color change of the silica gel to blue. After this initial reaction, the solution was relatively water insensitive, and the remaining water for complete hydrolysis was added in an acidified, alcoholic (i.e., 2-propanol) solution. The final stock solution, after filtration of the silica gel, was adjusted to 20 wgt.% solids in a 0.1M HCl, isopropanol solution. Further dilutions of the stock solution were made with anhydrous isopropanol. All coating solutions were usually utilized within a few days, but diluted samples had a shelf life in excess of one (1) month.

From a procedural perspective, it was found that a dried silica-to-water loading ratio of 5 provided sufficient adsorption capacity to ensure that there was a minimum of surface water on the gel. Lower loading ratios, where the silica gel surface appeared wet, resulted in the formation of a polymeric gel layer on the particle which both impeded further water diffusion and resulted in localized precipitation of aluminum-rich solids. Higher silica-to-water loading ratios were avoided as it was felt that the abrasive action associated with the mixing of the solids suspension would result in the formation of small particles which would be difficult to separate from the somewhat viscous stock solution.

The high refractive index stock solution was prepared in a similar manner starting with a 15:10:75 (molar) ratio of γ -glycidoxypolytrimethoxysilane, diethoxydiphenylsilane, and titanium (IV) isopropoxide.

Thin-Film Processing

Thin films were deposited on polycarbonate (Makrolon 3103) and silicon substrates by spin coating using a commercial photoresist spinner (Headway Research Corp.). As per the manufacturer's recommendation, the polycarbonate substrates were ultrasonically cleaned at room temperature in a bath of hexane, followed by distilled water, followed by isopropanol. The cleaned samples were finally blown dried which also served to remove remaining particulates. In a similar manner, the silicon substrates were cleaned in isopropanol and distilled water.

Film thickness was controlled by adjusting the solids content of the coating solution, as all films were processed at a spinner rate of 3000 rpm for 30 seconds. The deposited films were thermally cured in a convection oven operated at 125 °C for 20 minutes. Multilayered films were fabricated by pre-curing the

intermediate layers, as failure to do so resulted in the partial dissolving of the uncured polymer into the liquid coating solution. Samples prepared for subsequent AR performance characterization were coated on both sides of the substrate.

Analysis

The thickness and refractive index for single-layer films deposited on (100) silicon was determined by ellipsometry using a He-Ne laser source. An empirical correlation for coating solution solids concentration versus cured film thickness was determined for each coating formulation. This calculated data was used without correction during the processing of AR films on polycarbonate.

Reflectivity versus wavelength (380-780 nm) was measured on a Perkin-Elmer Lambda 9 spectrophotometer equipped with a planar reflectance accessory. The reflectance was reported as the % Reflectance as compared to a silver mirror standard (100% Reflectance).

Results

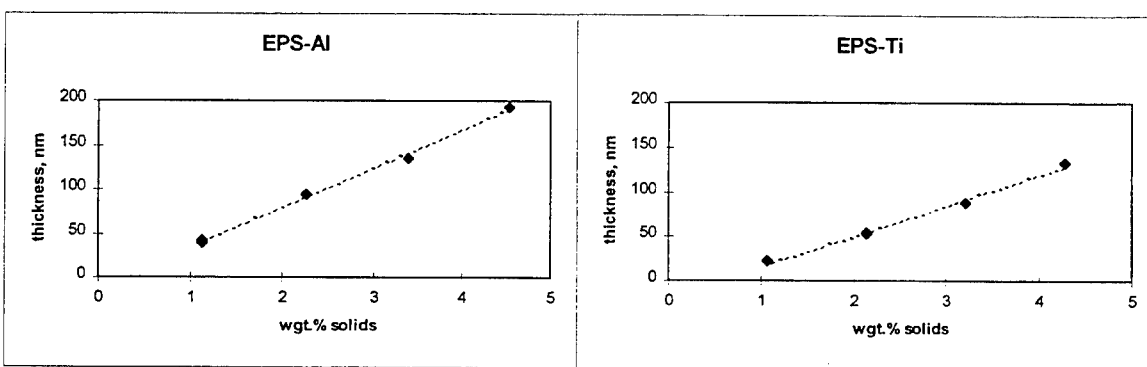
Antireflection Film Design

Based on ellipsometry measurements, the refractive index calculated at the He-Ne wavelength for the low index, EPS-Al, and high index, EPS-Ti, formulations were 1.48 ± 0.02 and 1.70 ± 0.05 , respectively. Figures 1 and 2 show the effect of coating solution solids content on the resulting cured film thickness for

Thickness (nm) as a function of coating solids content (wgt.%)

Figure 1

Figure 2



both formulations. All films were deposited by spin coating at 3000 rpm for 30 seconds. The deposited films were thermally cured at 125-130 °C for 30 minutes.

AR films were designed to achieve minimum reflectance at 510 nm. While the optical properties of the coating formulations studied did not lend themselves to the design of a zero reflectance system, the

basic guidelines used in the design of standard AR coatings served to define film thickness, and in the case of multilayered systems, placement of coating layers. Table 3 summarizes the design conditions for the

Table 3
Design Guides for AR Coatings
($\lambda_0 = 510$ nm)

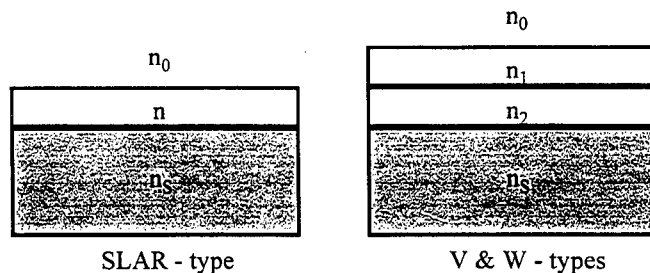
	$R_{\min} = 0$ Condition	Theoretical Coating(s) n_D	Optical Thickness	Theoretical Film Thickness (nm)
SLAR-Type	$n = \sqrt{n_0 n_s}$	$n = 1.259$	$\lambda/4$	101.3 nm (n)
V-Type	$\frac{n_2}{n_1} = \sqrt{\frac{n_s}{n_0}}$	$n_2 = 1.259 \cdot n_1$	$\lambda/4 - \lambda/4$	127.5/ n_1 nm (n_1) 127.5/ n_2 nm (n_2)
W-Type	N.A.	$n_2 = 1.259 \cdot n_1$	$\lambda/4 - \lambda/2$	127.5/ n_1 nm (n_1) 255.0/ n_2 nm (n_2)

where, $n_s = 1.586$ for polycarbonate, and $n_0 = 1.0$ for air.

standard single-layer $\lambda/4$ AR (SLAR) coating, the two-layer $\lambda/4$ - $\lambda/4$ V-type coating, and the broadband $\lambda/4$ - $\lambda/2$ W-type coating experimentally investigated. In order to fabricate test coupons, stock solutions of the EPS-Al and EPS-Ti formulations were diluted to the solids content dictated by film thickness targets. AR coatings were deposited on both sides of the polycarbonate substrate blank. Table 4 shows the design specifications for the fabricated AR films.

Table 4
Design Parameters for Fabricated AR Coatings
($\lambda_0 = 510$ nm)

	Refractive Index, n_D	Thickness, nm
SLAR-type	$n = 1.48$	86 (n)
V-type	$n_1 = 1.48$	86 (n_1)
	$n_2 = 1.70$	75 (n_2)
W-type	$n_1 = 1.48$	86 (n_1)
	$n_2 = 1.70$	150 (n_2)



Optical Performance of AR Films on Polycarbonate

The optical performance for the three AR film designs and the Control are summarized in Table 5, while the measured spectral reflectance within the visible wavelength range is graphically presented in

Table 5
Optical Performance Summary
Wavelength Range : 380 - 780 nm

% Reflection		
	Average	Median
Control (Makrolon 3103)	10.62 ± 0.89	10.83
SLAR-Coating	6.22 ± 0.79	6.09
V-Coating	4.70 ± 2.26	4.50
W-Coating	4.20 ± 1.20	3.99

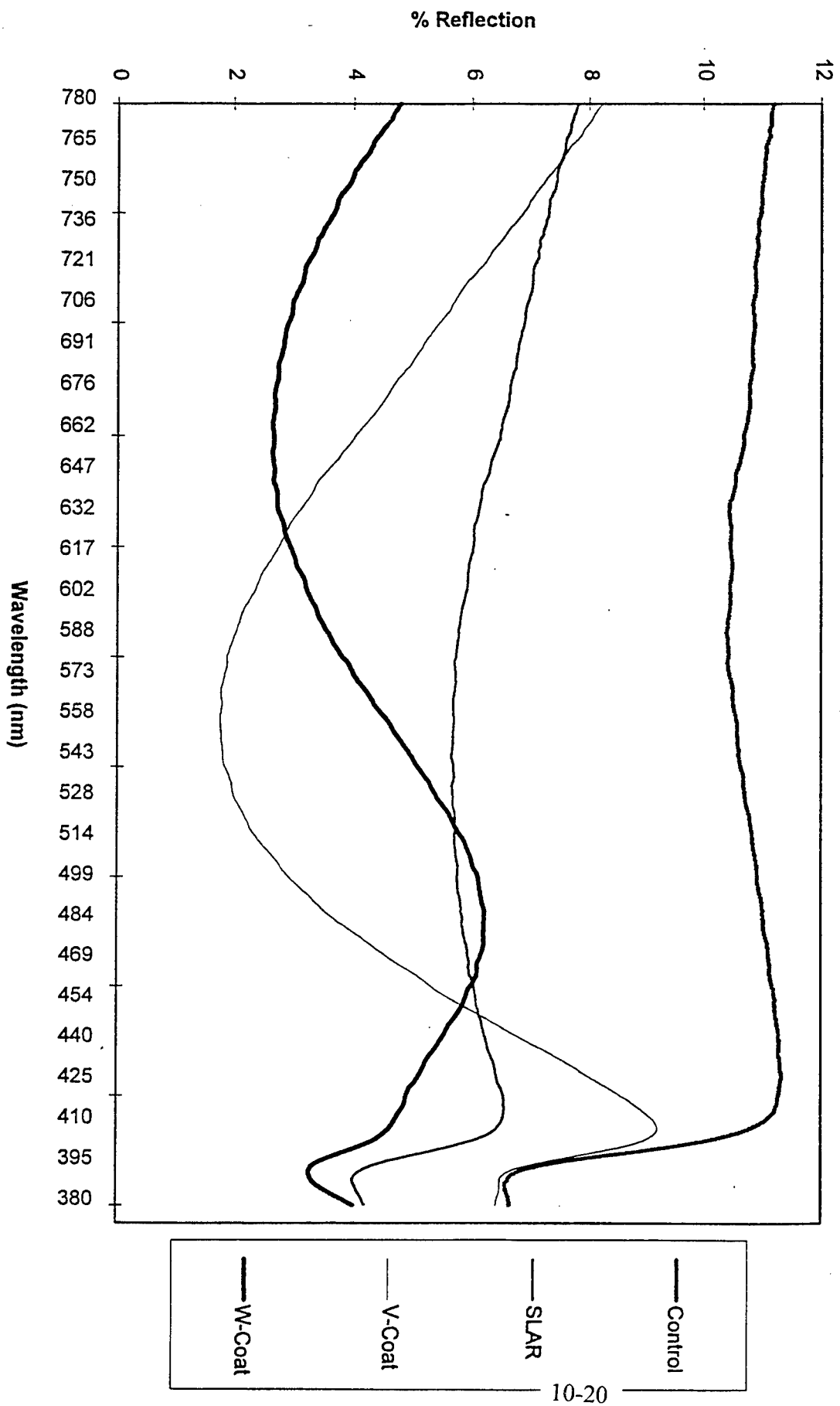
Figure 3. As noted in Figure 3, the minimum reflectance wavelength for both the SLAR and V-type coatings centers around 530-560 nm, as compared to the design specification of 510 nm. This small discrepancy probably results from the assumption that the film thickness would not be effected by the surface chemistry of the substrate. In general, the shape of the spectral reflectance curve is consistent with that expected from theory. Quantitatively, the resulting increase in optical transmission for the AR coated polycarbonate samples results in an optical performance similar to the lower refractive index polymethylmethacrylates and, as such provides a means to utilize the mechanically superior polycarbonates in critical optical applications.

Conclusion

AR coatings based on ORMOCer formulations were deposited at low temperature onto polymeric substrates. While the optical properties of the coatings were not optimized, the durability of the coatings were improved over that reported in the technical literature. It is notable that the same formulations have been used as abrasion resistant, protective coatings.

It is proposed to continue this effort by investigating alternative formulations for both high and low refractive index applications. In the case of high index materials, it may be possible to incorporate more “organic functionality” to raise the refractive index. Given that this film provides the interface between the substrate and low index material, surface hardness need not be a major concern. For the low index case, it is proposed to investigate nanocomposite processing schemes. Specifically, it may be possible to load the ORMOCer coating solution with nanoparticles of low-index solids (e.g., MgF_2). A multilayer of varying solids loading (i.e., a gradient) would also be of interest as a viable method to achieve broadband performance with a relatively thick protective film. The principal technological challenge to this approach is to find a processing technology suitable for the formation of metal fluoride nanoparticles.

Figure 3
Optical Performance of AR Coatings
(Substrate : Makrolon 3103)



ACKNOWLEDGEMENT

The author would like to acknowledge the support of the technical staff at Rome Laboratory Hanscom AFB, in particular, Michael Suscavage and Robert Andrews, for their work in material analysis and characterization.

Measurement of Free-Space THz Pulses via Long-Lifetime Photoconductors

X.-C. Zhang
Associate Professor
Department of Physics

Rensselaer Polytechnic Institute
Troy, NY

Final Report for:
Summer Research Extension Program

Sponsored by:
Air Force Office of Scientific Research
Bolling Air Force Base
Washington, DC

and

Rome Laboratory

December 1995

Measurement of Free-Space THz Pulses via Long-Lifetime Photoconductors

X.-C. Zhang

Associate Professor

Physics Department

Rensselaer Polytechnic Institute

Troy, NY 12180-3590

Abstract

We have tested photosensitivity, bandwidth, and device performance between regular GaAs and low-temperature grown GaAs as photoconductor for THz measurement. Antennas, based on commercially available GaAs as a photoconductor with a sub-nanosecond photocarrier lifetime, have been used to detect sub-picosecond free-space electromagnetic radiation (THz pulses). At low optical gating intensities ($\leq 1 \text{ mW}/100 \text{ }\mu\text{m}^2$) GaAs based antennas exhibit a higher responsivity and signal-to-noise ratio than typical antennas based on radiation-damaged silicon-on-sapphire.

Measurement of Free-Space THz Pulses via Long-Lifetime Photoconductors

X.-C. Zhang

Introduction

It is well known that sub-picosecond terahertz bandwidth (THz) electromagnetic (EM) pulses can be generated via femtosecond illumination of photoconductive emitters which have nanosecond photocarrier lifetimes [1,2]. This is because the far field radiation is proportional to the time derivative of the transient photocurrent. Thus, a step-like photocurrent produces a delta-like EM pulse in the far field. The duration of the radiated pulse is limited by the risetime of the induced photocurrent which may approach the duration of optical excitation. Correspondingly, the bandwidth of these pulses is typically several THz. Currently, time-resolved detection of THz waveforms is based on photoconductive sampling using materials which have extremely short, typically sub-picosecond, photocarrier lifetimes [3]. These materials include radiation-damaged silicon-on-sapphire (RD-SOS), low temperature grown (LT-) semiconductors, and host:precipitate GaAs:As [4]. With this technique, shorter photocarrier lifetimes provide improved temporal resolution and signal-to-noise ratio (SNR). However, the high defect density in these materials, which reduces photocarrier lifetime, also results in reduced mobility which limits the detection photocurrent.

Recently, an ultrafast measurement in an integrated circuit, using a long photocarrier lifetime material, in conjunction with a differential coupling technique, was demonstrated [5]. Later, an integrated inductive loop, containing an interdigitized photoconductive gate embedded in semi-insulating (SI) GaAs, was used for picosecond photoconductive sampling [6]. Similarly, it is possible to extend the concept of using long photocarrier lifetime materials to measure free space THz waveforms. In this paper, we present the ultrafast measurement and analysis of sub-picosecond free-space THz pulses using long lifetime (≈ 1 ns) photoconducting antennas. We then directly correlate these results with those obtained using typical ultrashort lifetime photoconducting antennas. At low optical gating intensities (≤ 1

mW/100 μm^2) GaAs based antennas exhibit a higher responsivity and signal-to-noise ratio than typical antennas based on radiation-damaged silicon-on-sapphire. Our power scaling results indicate the potential use of these devices with amplified optical pulses, large-aperture detectors, and THz antenna array.

Experimental Set up

The experimental setup used for time-resolved photoconductive sampling of free-space THz pulses has been described elsewhere [7]. The THz radiation source was an unbiased GaAs wafer. With the small gap antennas, we used a microscope objective to focus the optical gating beam and a silicon lens to focus THz beam onto the antenna.

Several THz receivers, co-planar transmission lines, were fabricated on undamaged SI-GaAs wafers. The photocarrier lifetime of this material is about several hundred picoseconds as measured by time-resolved photomodulated reflectance. The gaps between the 1 cm long antenna electrodes varied from 5 μm to 4 mm.

Experimental Results

Neglecting dipole resonance effects, the signal photocurrent $I(\tau)$ from a photoconductive sampling measurement is the correlation of the transient bias field $E(t)$ and the gating photocarrier density $n(t)$

$$I(\tau) = \int_{-\infty}^{\infty} dt E(t) n(t + \tau), \quad (1)$$

where τ is the relative time delay between THz signal and optical gating pulse.

Fig. 1 is a plot of the temporal photocurrent $I(\tau)$ from a regular GaAs based photoconducting antenna with a 5- μm electrode gap. The average optical power is 38 mW. The laser spot on the detector is about 100 μm^2 . With an unbiased GaAs wafer emitter and mW optical power, photocurrent is in nA range. The rise/fall time of the observed peak is 0.8/0.7 ps, respectively. Even though the GaAs photoconductor has a sub-nanosecond carrier lifetime, the photocurrent $I(\tau)$ yields a well resolved sub-picosecond signal. From Eqn. (1), the rising edge in the detected

photocurrent corresponds to the integration of the positive amplitude half-cycle of THz field, and the maximum value of $I(\tau)$ is proportional to the total area of positive field. The time between the rising point and the maximum of the photocurrent is the half-cycle time of the THz signal. The falling edge indicates the integration of the negative half-cycle of the THz field.

In an antenna where the photocarrier lifetime is much less than the duration of the THz signal, the time dependent carrier density $n(t)$ can be treated as a delta function $\delta(t)$, which from Eqn. 1 gives $I(\tau) = E(\tau)$. However, if the photoconductor has a long carrier lifetime with an ultrafast risetime, such as the GaAs we used here, to first order, the temporal response of $n(t)$ can be treated as a step function $\theta(t)$. Therefore the derivative of Equation (1) has the form of

$$\frac{d}{d\tau} I(\tau) = \int_{-\infty}^{\infty} dt E(t) \frac{d}{d\tau} \theta(t + \tau) = \int_{-\infty}^{\infty} dt E(t) \delta(t + \tau) = E(\tau), \quad (2)$$

Equation (2) provides a simple relation for the decorrelation of the step-like photocurrent from the radiated field, where the derivative of the measured photocurrent is proportional to the THz waveform. Fig. 2 is a plot of the numerical derivative of the signal shown Fig. 1. As expected, this waveform gives the temporal waveform of THz signal which is similar to the waveform of the photocurrent obtained using an ultrafast RD-SOS detector. Further, the spectrum bandwidth of the measured signal extends beyond 1 THz.

To investigate the performance of SI-GaAs based long-lifetime photoconductive sampling on gating intensity, we have measured THz signals while varying the optical intensity incident on the photoconductor from less than $1 \mu\text{W}/100 \mu\text{m}^2$ to $38 \text{ mW}/100 \mu\text{m}^2$. Fig. 3 and Fig. 4 show the peak photocurrent signal versus average gating power for GaAs and RD-SOS based photoconducting antennas ($100\text{-}\mu\text{m}$ dipole detectors), respectively. Throughout this range, the shape of the THz waveforms from both detectors remained the same. At moderate gating intensities, RD-SOS devices offer better responsivity and SNR characteristics (not shown). However, when using low gating intensities, SI-GaAs based antennas are clearly the better choice, and a SNR greater than 100 has been achieved with a gate

intensity less than $1 \mu\text{W}/100 \mu\text{m}^2$. Further, at $330 \text{ nW}/100 \mu\text{m}^2$ intensity levels, the THz signal could not be observed using a RD-SOS based antenna, while it was clearly resolved with the SI-GaAs device. The improved responsivity of SI-GaAs over RD-SOS at low gate light intensities is due to its higher photocarrier mobility and strong optical absorption near 800 nm . For GaAs based antennas, the photocurrent signal begins to saturate above $1 \text{ mW}/100 \mu\text{m}^2$ gating intensity, whereas a similar RD-SOS based antenna exhibits no saturation up to $38 \text{ mW}/100 \mu\text{m}^2$. It is not completely understood why the GaAs detector shows the photocurrent saturation in Fig. 3. One possible reason of the saturation is the field screening by photocarriers. Our measured of THz field and the calculated Coulomb's field in GaAs are in the same order when a light intensity of $0.5 \text{ mW}/100 \mu\text{m}^2$ is used. Once the screening field in the photoconductor is comparable with the applied THz field, the photocurrent may saturate. RD-SOS based antennas exhibit saturation until much higher intensities ($> 50 \text{ mW}/100 \mu\text{m}^2$) due to a reduced carrier concentration resulting from the long absorption length of silicon near 800 nm .

SI-GaAs offers distinct advantages in devices where the gating beam is distributed over a large area. Two types of devices we are investigating require a distributed gating beam: directional large-aperture antennas and synchronously gated antenna arrays. The high photosensitivity of SI-GaAs permits the fabrication of device arrays and large-aperture detectors with improved responsivity. Synchronously gated antenna arrays will be used in time-resolved THz imaging applications [8]. Large-aperture photoconducting antennas will be used as directional THz detectors for intense THz field measurements based on amplified optical pulse.

As a proof of concept, we measured a THz waveform using a large-aperture SI-GaAs based photoconducting detector consisting of two parallel electrodes which are separated by 2 mm . During this measurement, both the objective used to focus the gating beam and the silicon lens used to focus the THz beam onto the antenna were removed from the system. Fig. 5 shows the results of this measurement. Even with no focusing elements, which reduces both the THz field strength and detector

response by more than four orders of magnitude, the signal is still clearly resolvable.

Conclusion

We report the results of photoconductive sampling of THz pulses with SI-GaAs based long carrier lifetime antennas. At low optical gating intensities, SI-GaAs exhibits a higher responsivity and SNR than compared to RD-SOS. Our results indicate that antennas based on regular GaAs is an alternative device for directional large-aperture detector structures and THz imaging applications using detector arrays. Another parallel project we are working now is to compare detection performance between regular GaAs and LT-GaAs photoconductors, those results will be reported in a following paper.

Acknowledgment

We would like to thank M.N. Alexander, J.J. Larkin, M.T. Harris, B.S. Ahern, D. Bliss, and D.W. Weybume for their support during the 1995 AFOSR Faculty Research Extension Program.

References:

- [1] D. Krokel, D. Grischkowsky, and M.B. Ketchen, Appl. Phys. Lett., **54**, 1046 (1989).
- [2] J.T. Darrow, B.B. Hu, X.-C. Zhang and D.H. Auston, Opt. Lett. **15**, 323 (1990).
- [3] For example, see M. van Exter, C. Fattinger, and D. Grischkowski, Appl. Phys. Lett., **55**, 337 (1989).
- [4] A.C. Warren, N. Katzenellenbogen, D. Grischkowsky, J.M. Woodall, M.R. Melloch, and N. Otsuka, Appl. Phys. Lett., **58**, 1512 (1991).
- [5] J. Paslaski and A. Yariv, Appl. Phys. Lett., **55**, 1744 (1989).
- [6] A.C. Davidson, F. W. Wise, and R.C. Compton, Appl. Phys. Lett., **66**, 2259 (1995).
- [7] X.-C. Zhang, Y. Jin, K. Yang, and L. J. Schowalter, Phys. Rev. Lett. **69**, 2303 (1992).
- [8] B. B. Hu and M. C. Nuss, Technical Digest of the Ultrafast Electronics and Optoelectronics Topical Meeting, Optical Society of America, Technical Digest Series **13**, PD2-2 (1995).

Figure Captions:

- Fig. 1 Sampled photocurrent signal using a long lifetime regular GaAs based antenna with a 5 μm photoconducting gap. The THz emitter is an unbiased SI-GaAs wafer.
- Fig. 2 THz signal: numerical derivative of the photocurrent signal in Fig. 1.
- Fig. 3 Peak photocurrent versus gating intensity for GaAs based 5 μm photoconducting antenna.
- Fig. 4 Peak photocurrent versus gating intensity for RD-SOS based 5 μm photoconducting antenna.
- Fig. 5 Sampled photocurrent signal using a long lifetime GaAs antenna with a 2-mm photoconducting gap. No focusing elements for either the THz beam or the gating beam were used.

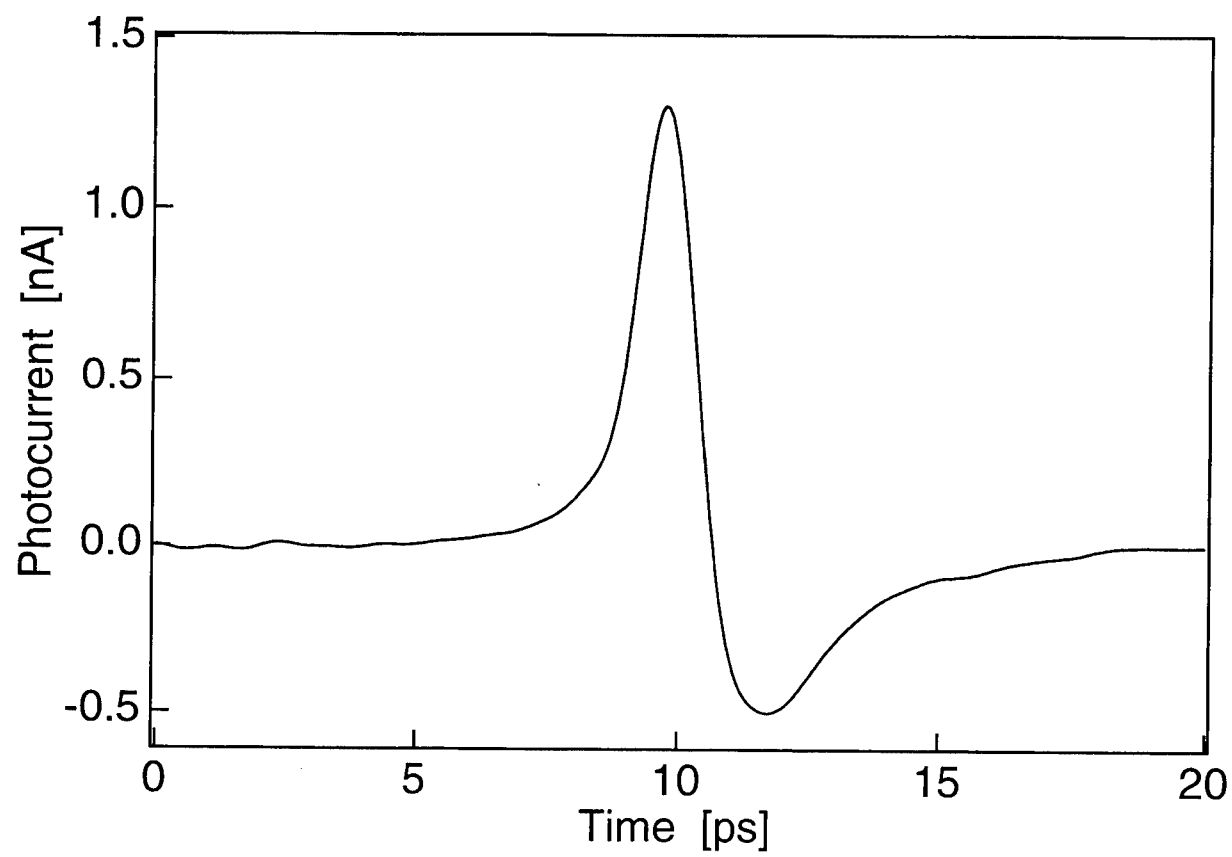


Fig. 1

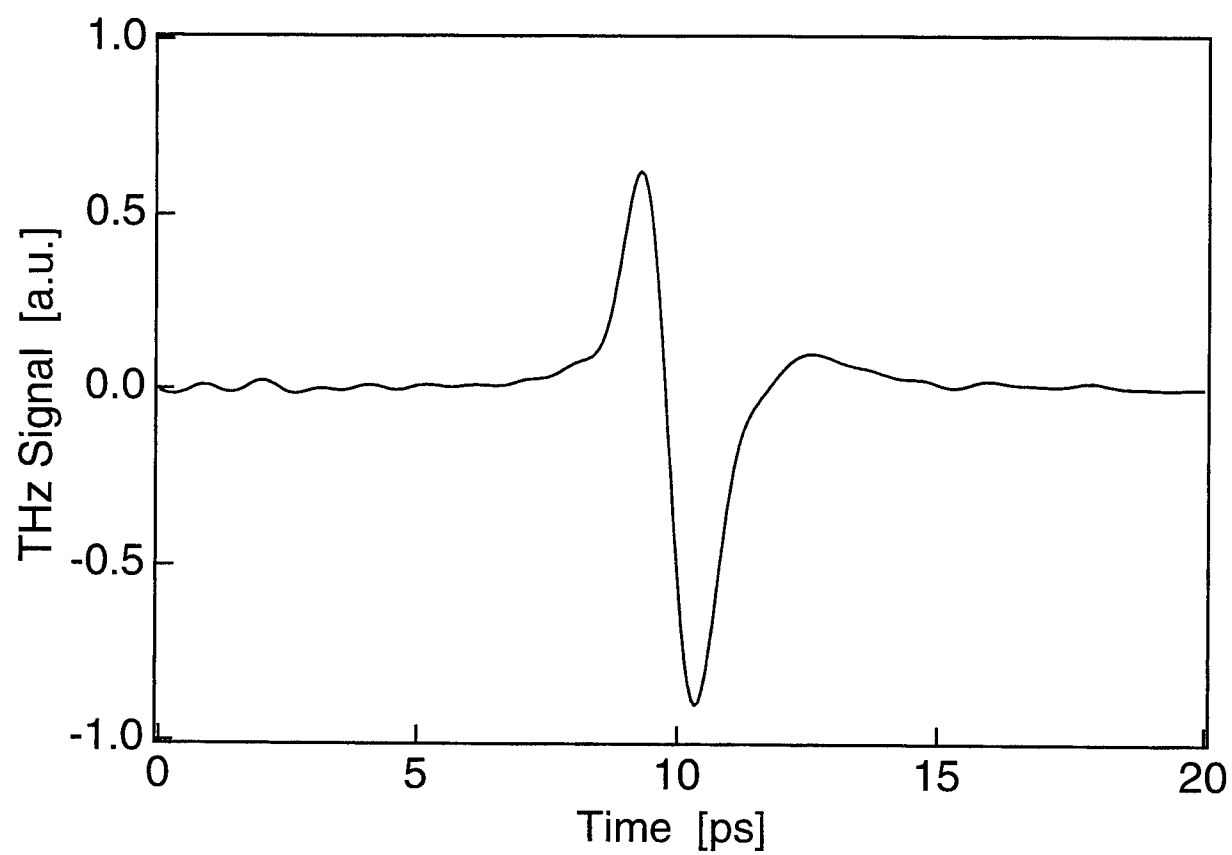


Fig. 2

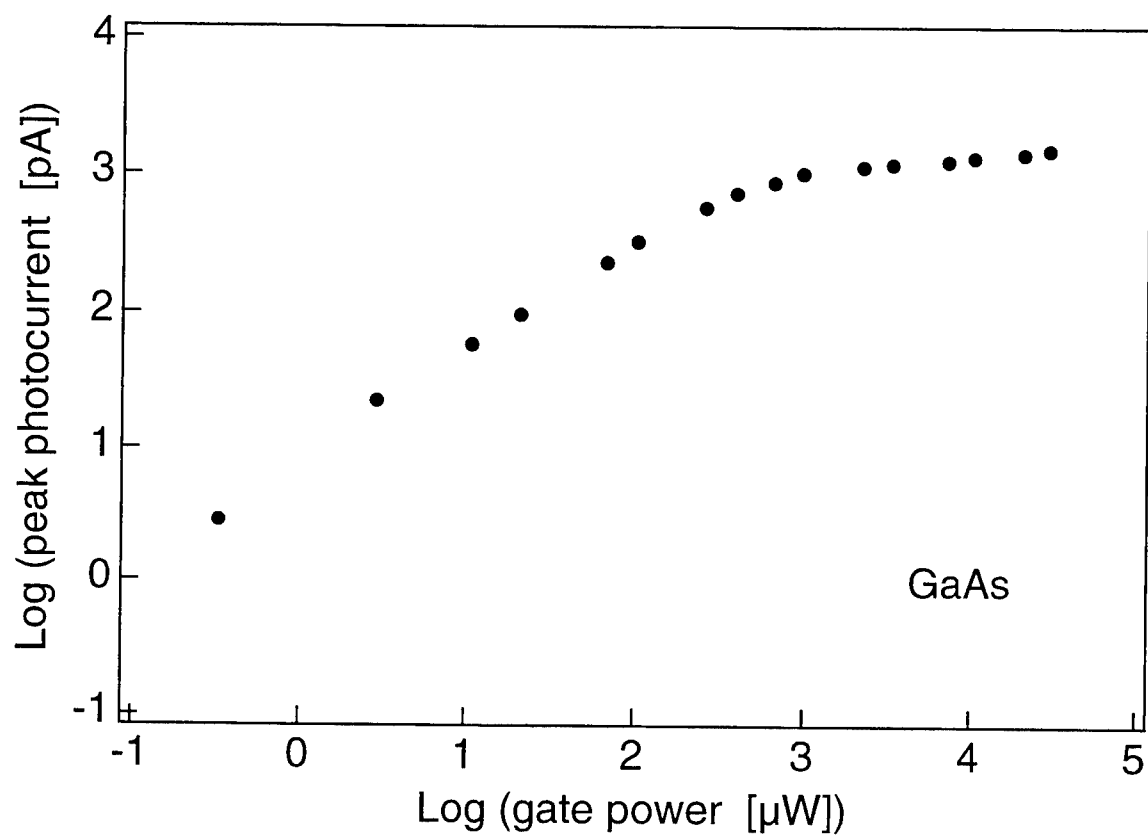


Fig. 3

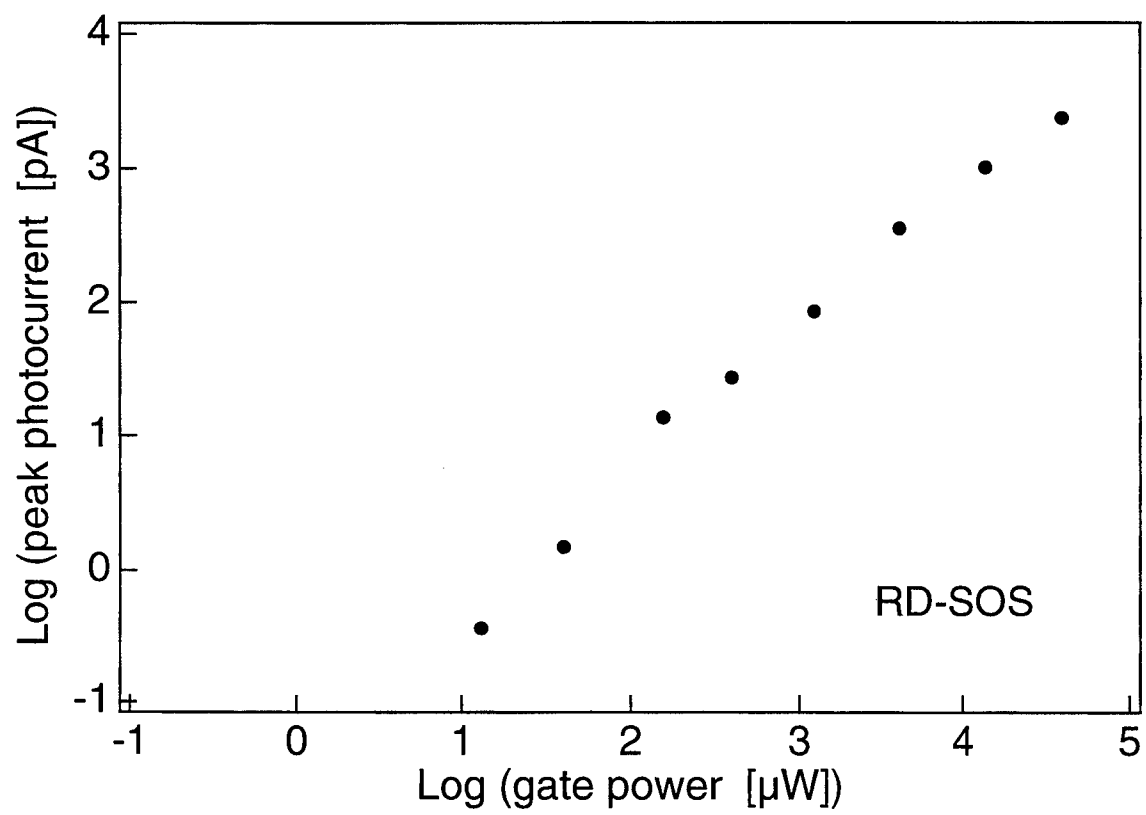


Fig. 4

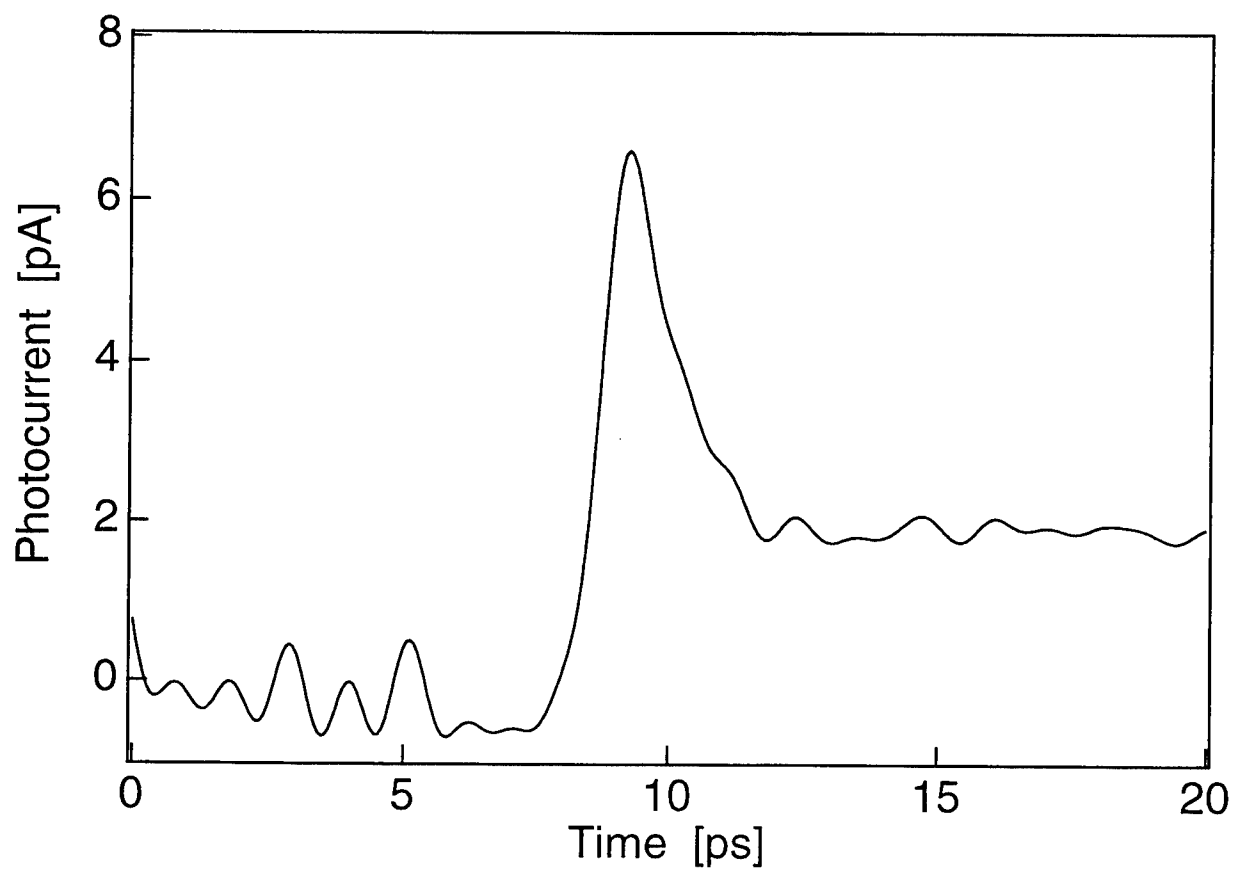


Fig. 5